

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 14:59:04 ; Search time 338 Seconds  
(without alignments)  
9462.933 Million cell updates/sec

Title: US-09-937-521-15  
Perfect score: 2049  
Sequence: 1 ttcaagtgctactactatgg.....ttgaagtgctactactatg 2049

Scoring table:  
OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 1085931 seqs, 780495707 residues

Word size: 20

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: Published Applications\_NA:\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	29	1.4	1194	10	US-09-739-907-49
2	29	1.4	1222	10	US-09-739-907-31
3	22	1.1	4022	9	US-09-964-899-18
4	20	1.0	504	9	US-09-918-995-20201

## ALIGNMENTS

RESULT 1  
US-09-739-907-49  
Sequence 49, Application US/09739907  
Patent No. US20010012889A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 36 Human Secreted Proteins  
FILE REFERENCE: P2022P1  
CURRENT APPLICATION NUMBER: US/09/739, 907  
CURRENT FILING DATE: 2000-12-20

```
;; PRIOR APPLICATION NUMBER: 09/348,457
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: 60/070,567
;; PRIOR FILING DATE: 1998-01-07
;; PRIOR APPLICATION NUMBER: 60/070,692
;; PRIOR FILING DATE: 1998-01-07
;; PRIOR APPLICATION NUMBER: 60/070,704
;; PRIOR FILING DATE: 1998-01-07
;; PRIOR APPLICATION NUMBER: 60/070,658
;; PRIOR FILING DATE: 1998-01-07
;; NUMBER OF SEQ ID NOS: 196
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 49
;; LENGTH: 1194
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-739-907-49
```

Query Match 1.4%; Score 29; DB 10; Length 1194;  
Best Local Similarity 100.0%; Pred. No. 3.1e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CAGATATCAATTGATGGCTATGCGCAA 81  
|||||  
Db 478 CAGATATCAATTGATGGCTATGCGCAA 506

RESULT 2  
US-09-739-907-31  
Sequence 31, Application US/09739907  
Patent No. US20010012889A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 36 Human Secreted Proteins  
FILE REFERENCE: P2022P1  
CURRENT APPLICATION NUMBER: US/09/739, 907  
CURRENT FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 09/348,457  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: 60/070,567  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: 60/070,692  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: 60/070,704  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: 60/070,658  
PRIOR FILING DATE: 1998-01-07  
NUMBER OF SEQ ID NOS: 196  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 31  
LENGTH: 1222  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-739-907-31

Query Match 1.4%; Score 29; DB 10; Length 1222;  
Best Local Similarity 100.0%; Pred. No. 3.1e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CAGATATCAATTGATGGCTATGCGCAA 81  
|||||  
Db 501 CAGATATCAATTGATGGCTATGCGCAA 529

RESULT 3  
US-09-964-899-18  
Sequence 18, Application US/09964899  
Patent No. US20020174446A1  
GENERAL INFORMATION:  
APPLICANT: Cohen, Dalja et al.  
TITLE OF INVENTION: Identification of Genes Involved in  
TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster  
FILE REFERENCE: 4-31612 A

;; CURRENT APPLICATION NUMBER: US/09/964,899  
;; CURRENT FILING DATE: 2001-09-27  
;; PRIOR APPLICATION NUMBER: 60/236,893  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/298,309  
;; PRIOR FILING DATE: 2001-06-14  
;; NUMBER OF SEQ ID NOS: 53  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 18  
;; LENGTH: 4022  
;; TYPE: DNA  
;; ORGANISM: Homo Sapien  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(4022)  
;; OTHER INFORMATION: n - A,T,C or G  
US-09-964-899-18

Query Match 1.1%; Score 22; DB 9; Length 4022;  
Best Local Similarity 100.0%; Pred. No. 0.29;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAGGCCAGGAGCTGTATG 943  
DB 319 CAGAGGCCAGGAGCTGTATG 340

RESULT 4  
US-09-918-995-20201/c  
;; Sequence 20201, Application US/09918995  
;; Publication No. US20030073623A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc.  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
;; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
;; FILE REFERENCE: 20411-756  
;; CURRENT APPLICATION NUMBER: US/09/918,995  
;; CURRENT FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: US/09/235,076  
;; PRIOR FILING DATE: 1999-01-20  
;; NUMBER OF SEQ ID NOS: 38054  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO: 20201  
;; LENGTH: 504  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(504)  
;; OTHER INFORMATION: n - A,T,C or G  
US-09-918-995-20201

Query Match 1.0%; Score 20; DB 9; Length 504;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1056 TGCCCTGGGCTACACTTTTG 1075  
DB 78 TGCCCTGGGCTACACTTTTG 59

Search completed: July 5, 2003, 17:40:52  
Job time : 339 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 12:55:04 ; Search time 100 Seconds  
(without alignments)  
6283.807 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgtactactctg.....ttgaagttgtactactag 2049

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents-NA:\*

1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PTCUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Dackfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	1.1	4020	4	US-09-050-159-130
2	22	1.1	4024	4	US-09-162-484-18
3	21	1.0	3942	4	US-09-162-484-19

## ALIGNMENTS

RESULT 1  
US-09-050-159-130  
; Sequence 130, Application US/09050159A  
; Patent No. 6197505  
; GENERAL INFORMATION:  
; APPLICANT: No. 6197505berg, Leif T  
; APPLICANT: Andersson, Maria K  
; APPLICANT: Linstrom, Per H  
; TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND  
; FILE REFERENCE: 1248/ID042  
; CURRENT APPLICATION NUMBER: US/09/050,159A  
; CURRENT FILING DATE: 1998-03-27  
; EARLIER APPLICATION NUMBER: 60/042,930  
; EARLIER FILING DATE: 1987-04-03  
; NUMBER OF SEQ ID NOS: 133  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 130  
; LENGTH: 4020

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Angiotensin I converting enzyme mRNA  
US-09-050-159-130

Query Match 1.1%; Score 22; DB 4; Length 4020;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAAGCCCAAGAGCTGTATG 943  
DB 317 CAGAAGCCCAAGAGCTGTATG 338

RESULT 2  
US-09-162-484-18  
; Sequence 18, Application US/09162484  
; Patent No. 6248724  
; GENERAL INFORMATION:  
; APPLICANT: Phillips, M. Ian  
; APPLICANT: Mohuczy, Dagmara  
; TITLE OF INVENTION: ANGIOGENESIS CONVERTING ENZYME MRNA AND METHODS OF USE  
; FILE REFERENCE: UFLA:087/UFLA087P  
; CURRENT APPLICATION NUMBER: US/09/162,484  
; CURRENT FILING DATE: 1998-09-25  
; EARLIER APPLICATION NUMBER: 60/059,661  
; EARLIER FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 4024  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-162-484-18

Query Match 1.1%; Score 22; DB 4; Length 4024;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 CAGAAGCCCAAGAGCTGTATG 943  
DB 317 CAGAAGCCCAAGAGCTGTATG 338

RESULT 3  
US-09-162-484-19  
; Sequence 19, Application US/09162484  
; Patent No. 6248724  
; GENERAL INFORMATION:  
; APPLICANT: Phillips, M. Ian  
; APPLICANT: Mohuczy, Dagmara  
; TITLE OF INVENTION: ANGIOGENESIS CONVERTING ENZYME MRNA AND METHODS OF USE  
; FILE REFERENCE: UFLA:087/UFLA087P  
; CURRENT APPLICATION NUMBER: US/09/162,484  
; CURRENT FILING DATE: 1998-09-25  
; EARLIER APPLICATION NUMBER: 60/059,661  
; EARLIER FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 3942  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-09-162-484-19

Query Match 1.0%; Score 21; DB 4; Length 3942;  
Best Local Similarity 100.0%; Pred. No. 0.94;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 923 AGAAGCCCAAGAGCTGTATG 943

Mon Jul 7 09:49:08 2003

us-09-937-521-15.oli20.rni

Page 2

Db 314 AGAAGGCCAAGAGAGCTGTATG 334

Search completed: July 5, 2003, 15:50:30  
Job time : 100 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 12:49:14 ; Search time 2980 Seconds  
(without alignments)  
11135.756 Million cell updates/sec

Title: US-09-937-521-15  
Perfect score: 2049  
Sequence: 1 ttcaatgagctactactatg.....ttgaagtgtcactactatg 2049

Scoring table: ORIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 20

Total number of hits satisfying chosen parameters: 44

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST: \*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estnu:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	655	32.0	700	10	BB356080 BB356080
2	536	26.2	896	12	BP237061 602027402
3	157	7.7	451	9	AA920146 v92h04.t
4	155	7.6	674	17	AZ018768 RPCI-23-3
5	125	6.1	834	17	BH049419 RPCI-24-3
6	112	5.5	269	9	AV284970 AV284970

7	100	4.9	292	17	AZ907075
8	97	4.7	451	17	AZ879271
9	85	4.1	698	10	BB327376
10	83	4.1	281	10	BB356081
11	77	3.8	551	17	AZ727531
12	77	3.8	701	17	AZ562461
13	75	3.7	277	10	BB242940
14	73	3.6	312	10	BB215231
15	70	3.4	246	9	AV160656
16	53	2.6	436	14	BQ193951
17	50	2.4	789	10	BE531642
18	47	2.3	335	12	BG209882
19	47	2.3	436	10	BB817265
20	47	2.3	484	12	BG289059
21	47	2.3	636	12	BG773189
22	47	2.3	672	9	AA913512
23	47	2.3	722	13	B1553338
24	47	2.3	813	10	BE387191
25	31	1.5	534	12	BG702763
26	31	1.5	806	12	BE971187
27	31	1.5	936	13	B1560354
28	29	1.4	503	9	A1229932
29	24	1.2	742	13	B1553673
30	23	1.1	581	17	AQ976049
31	23	1.1	384	9	AA373225
32	22	1.1	450	9	A1365378
33	21	1.0	1501	9	AL692789
34	20	1.0	220	17	AZ656375
35	20	1.0	264	9	AV210530
36	20	1.0	296	10	BB551563
37	20	1.0	299	10	BB556696
38	20	1.0	516	17	AZ386497
39	20	1.0	540	17	AQ696374
40	20	1.0	560	17	AZ794922
41	20	1.0	648	10	BB612587
42	20	1.0	795	12	BG481760
43	20	1.0	822	17	BH124617
44	20	1.0	1018	17	CNS07928

#### ALIGNMENTS

RESULT 1	BB356080	700 bp	mRNA	linear	EST 24-OCT-2001
LOCUS	BB356080	RIKEN full-length enriched, adult male corpus striatum Mus			
DEFINITION	BB356080	musculus cDNA clone C030011021 3', mRNA sequence.			
ACCESSION	BB356080	GI:16405735			
VERSION	BB356080.2	GI:16405735			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 700)				
	Arakawa,T., Carrincci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Komuro,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takehashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.				
	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)				
	Unpublished (2001)				
	On Jul 12, 2000 this sequence version replaced gi:9067908.				
TITLE	LABORATORY FOR GENOME EXPLORATION RESEARCH GROUP, RIKEN GENOMIC SCIENCES CENTER(GSC), YOKOHAMA INSTITUTE				
JOURNAL	The Institute of Physical and Chemical Research (RIKEN)				
COMMENT	1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan				
	Tel: 81-45-503-9222				
	Fax: 81-45-503-9216				



/clone\_lib="NCI\_CGAP\_Li9"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: Liver; Vector: pCMV-Sport6; Site: 1; NotI:  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP library."  
BASE COUNT 236 a 223 c 229 g 208 t  
ORIGIN

Query Match 26.2%; Score 536; DB 12; Length 896;  
Best Local Similarity 100.0%; Pred. No. 1.7e-270; Indels 0; Gaps 0;  
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 644 TTGTGATAGTACATATGAGTATGCGCTATGCTTTTGGACGAAAGAAAG 703  
DB 87 TTGTGATAGTACATATGAGTATGCGCTATGCTTTTGGACGAAAGAAAG 146  
QY 704 GCTATGCTGCGAGAGGAGCGCTTTGACGAGCTTTGCTCAATCTCGAGAGC 763  
DB 147 GCTATGCTGCGAGAGGAGCGCTTTGACGAGCTTTGCTCAATCTCGAGAGC 206  
QY 764 TGTACCCCAACATCTTGGCCCGATGCTGTCAACACAGGAGGCTTTGACACGCA 823  
DB 207 TGTACCCCAACATCTTGGCCCGATGCTGTCAACACAGGAGGCTTTGACACGCA 266  
QY 824 AGAGCAGCTGTCCCAAGGAGGAGCGCTTACATGTCATGCGAGGAGCTGACAGACA 883  
DB 267 AGAGCAGCTGTCCCAAGGAGGAGCGCTTACATGTCATGCGAGGAGCTGACAGACA 326  
QY 884 TGTGAGAGACACACATTTATGAGAGGATCATCTATCAGAAAGCCAGAGCTGTATG 943  
DB 327 TGTGAGAGACACACATTTATGAGAGGATCATCTATCAGAAAGCCAGAGCTGTATG 386  
QY 944 CCTGTGCTCCAGAGGAGGAGCGCGGCTTGTGAGCTGACCATGCTGTGAACATGA 1003  
DB 387 CCTGTGCTCCAGAGGAGGAGCGCGGCTTGTGAGCTGACCATGCTGTGAACATGA 446  
QY 1004 CAGATGTGAGCGTCCAGCTCAATGCCACACACAGTGAAGAGGTGAAACCTCCCTGG 1063  
DB 447 CAGATGTGAGCGTCCAGCTCAATGCCACACACAGTGAAGAGGTGAAACCTCCCTGG 506  
QY 1064 GCTACAGCTTTTGGCCGAGGACCAATTTATGAGAGGCTTGGGCGCTCAATTTACACAGGAA 1123  
DB 507 GCTACAGCTTTTGGCCGAGGACCAATTTATGAGAGGCTTGGGCGCTCAATTTACACAGGAA 566  
QY 1124 CTACGGAAGGGGATCCATCTTGGACACTCTTGGGAGCCAGCTTTGGGAAACCA 1179  
DB 567 CTACGGAAGGGGATCCATCTTGGACACTCTTGGGAGCCAGCTTTGGGAAACCA 622

RESULT 3  
AA920146 451 bp mRNA linear EST 20-APR-1998  
LOCUS  
DEFINITION  
IMAGE:1299127 5', mRNA sequence.  
AA920146  
AA920146.1 GI:3066925  
EST.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 451)

REFERENCE  
AUTHORS  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.  
The MashU-HMI Mouse EST Project  
Unpublished (1996)

TITLE  
JOURNAL  
COMMENT  
Contact: Marra M/Mouse EST Project  
MashU-HMI Mouse EST Project  
Washington University School of Medicine

*[Handwritten signature]*

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:680175  
Seq primer: -28ml3 rev1 ET from Amersham.

## FEATURES

location/Qualifiers  
1..451  
/organism="Mus musculus"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1299127"  
/clone\_lib="Stratagene mouse lung 937302"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: plasmid; SK-; Site: 1; EcoRI  
Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. 6-8 month old female lung and 1.5 year old male lung  
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP  
XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'  
adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3' "

## BASE COUNT

144 a 58 c 129 g 120 t

Query Match 7.7%; Score 157; DB 9; Length 451;  
Best Local Similarity 100.0%; Pred. No. 6.3e-71;  
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 CAATGCAACATATACGATGCGATATTCAGATACCTTACCTGGATCTACACAAATAG 1952  
DB 1 CAATGCAACATATACGATGCGATATTCAGATACCTTACCTGGATCTACACAAATAG 60  
QY 1953 ATATTTGACACACATGGAAGAGCAAGCTTGAACCCGCTGTACTAGCATTTGA 2012  
DB 61 ATATTTGACACACATGGAAGAGCAAGCTTGAACCCGCTGTACTAGCATTTGA 120  
QY 2013 AGCAATTTCTTCTTTTGAAGTTGCTACTACTAG 2049  
DB 121 AGCAATTTCTTCTTTTGAAGTTGCTACTACTAG 157

## RESULT 4

AZ018768/c 674 bp DNA linear GSS 25-FEB-2000  
LOCUS  
DEFINITION  
IMAGE:1299127 5', DNA sequence.  
AZ018768  
AZ018768.1 GI:7094152  
GSS.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 674)

REFERENCE  
AUTHORS  
Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akiret  
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P.  
and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)

TITLE  
JOURNAL  
COMMENT  
Other GSSs: RPCI-23-334M15.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong







Db 239 TCCTTTGAAGTGTCTACTACTAG 263

RESULT 10  
LOCUS BB356081

DEFINITION BB356081 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030011022 3', mRNA sequence.

ACCESSION BB356081

VERSION BB356081.1 GI:9067909

KEYWORDS EST.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Konno H., Aizawa K., Akahira S., Akiyama J., Arakawa T., Carninci P., Endo T., Fukuda S., Fukunishi Y., Hara A., Hayatsu N., Hirozane T., Horii F., Ishii Y., Ishikawa J., Ishikawa T., Itoh M., Iizawa M., Kadota K., Kagawa I., Kai C., Kawai J., Kikuchi N., Kiyosawa H., Kojima Y., Kondo S., Koya S., Kurihara C., Kusakabe M., Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y., Ono T., Owa C., Saito H., Sakai C., Sato K., Shibata K., Shibata Y., Shigemoto Y., Shinagawa A., Shiraki T., Sogabe Y., Sugahara Y., Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomimaga N., Toyota T., Tsunoda Y., Watanabe A., Watanabe S., Yamamura T., Yamanaoka I., Yano R., Yasunishi A., Yokota T., Yoshida K., Yoshiki A., Yoshino M., Muramatsu M. and Hayashizaki Y.

TITLE RIKEN Mouse ESTs (Konno H., et al.)

JOURNAL Unpublished (2000)

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
url: http://genome.gsc.riken.go.jp/  
Carninci P., Nishiyama Y., Westover A., Itoh M., Nagaoka S., Sasaki N., Okazaki Y., Muramatsu M. and Hayashizaki Y.  
Thermostabilization and thermoinactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh M., Kitsuina T., Akiyama J., Shibata K., Iizawa M., Kawai J., Tomaru Y., Carninci P., Shibata Y., Ozawa Y., Muramatsu M., Okazaki Y. and Hayashizaki Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci P. and Hayashizaki Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
source

location/Qualifiers

1..281

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="C030011022"

/clone\_1lb="RIKEN full-length enriched, adult male corpus striatum"

/sex="male"

/tissue\_type="corpus striatum"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'

BASE COUNT 81 a 72 c 55 g 73 t

ORIGIN

Query Match 4.1%; Score 83; DB 10; Length 281;  
Best Local Similarity 100.0%; Pred. No. 5.1e-32;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 608 CCATCCACCCCGTGAGCATGACATAGCAACACTTTGTGAATAGTACATATGGGCT 667  
|||||  
Db 199 CCATCCACCCCGTGAGCATGACATAGCAACACTTTGTGAATAGTACATATGGGCT 258  
|||||

Oy 668 ATGCGGCTTACCTTTTGAGCAA 690  
|||||  
Db 259 ATGCGGCTTACCTTTTGAGCAA 281  
|||||

RESULT 11  
LOCUS A2727531

DEFINITION RPCI-24-127B15.TJ RPCI-24 Mus musculus genomic clone RPCI-24-127B15, DNA sequence.

ACCESSION A2727531

VERSION A2727531.1 GI:12485027

KEYWORDS GSS.

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 551)  
Zhao S., Nierman W., Malek J., Shatsman S., Akiret B., Levins M., Tsegaye G., Geer K., Krol M., Shvartsbeyn A., Gebregorgis E., Russell D., de Jong P. and Fraser C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Other GSSs: RPCI-24-127B15.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@tigr.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tigr/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tigr/bac_ends/mouse/bac_end_intro.html)  
Plate: 127 row: B column: 15  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source

location/Qualifiers

1..551

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-127B15"

/clone\_1lb="RPCI-24"

/sex="male"

/cell\_type="Spleen/Brain"

/note="Vector: pPARBAC1; Site\_1: BamHI, Site\_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pPARBAC1 cloning vector at the



BamHI sites using MboI partially digested male C57BL/6J DNA.  
BASE COUNT 139 a 107 c 117 g 188 t  
ORIGIN

Query Match 3.8%; Score 77; DB 17; Length 551;  
Best Local Similarity 100.0%; Pred. No. 9.6e-29;  
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 718 CAGGACCGCTTGTAGCAGCGCTTGTCTCATCAATCTCGAGACGTGCACCAACATT 777  
|||||  
Db 82 CAGGACCGCTTGTAGCAGCGCTTGTCTCATCAATCTCGAGACGTGCACCAACATT 141

Qy 778 CTTGGCCCGCATTTGTGT 794  
|||||  
Db 142 CTTGGCCCGCATTTGTGT 158

RESULT 12  
A2562461 701 bp DNA linear GSS 20-NOV-2000  
LOCUS RPCI-23-207D7.TV RPCI-23 Mus musculus genomic clone RPCI-23-207D7,  
DEFINITION DNA sequence.  
ACCESSION A2562461 GI:11242281  
VERSION A2562461.1  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 701)  
Zhao, S., Nierman, W., Fajdlyum, T., Malek, J., Shatsman, S., Akınret,  
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P.  
and Fraser, C. M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-207D7.JU  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@edong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (inf@resgen.com). BAC end page:  
http://www.tigr.org/cdb/bac/ends/mouse/bac\_end\_intro.html  
Plate: 207 row: D column: 7  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..701  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-207D7"  
/clone\_1lb="RPCI-23"  
/sex="Female"  
/1ab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBAC3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 220 a 150 c 127 g 204 t  
ORIGIN

Query Match

3.8%; Score 77; DB 17; Length 701;

Best Local Similarity 100.0%; Pred. No. 1.1e-28;  
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1255 TGGCAACGAGATATTTGATGATTTGATTTAGCGTGGGCGCTGGCCATAGCTGCT 1314  
|||||  
Db 701 TGGCAACGAGATATTTGATGATTTGATTTAGCGTGGGCGCTGGCCATAGCTGCT 642

Qy 1315 ATCCCTGGGGAATTAC 1331  
|||||  
Db 641 ATCCCTGGGGAATTAC 625

RESULT 13  
BB242940 277 bp mRNA linear EST 06-JUL-2000  
LOCUS BB242940  
DEFINITION BB242940 RIKEN full-length enriched, 3 days neonate thymus Mus  
musculus cDNA clone A630094E02 3' similar to D17062 Human HepG2  
partial cDNA, clone hmd4c12m5, mRNA sequence.  
BB242940  
BB242940.1 GI:8935686  
EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 277)  
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci,  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Oho, T., Owa, C., Saito, H., Sakai, C., Saito, R., Shibata, K., Shibata,  
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,  
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,  
T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamaguchi, I.,  
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,  
M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Kono, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gscl.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki,  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermolabile and thermostable restriction enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.ritc.riken.go.jp) for  
further details.

FEATURES  
source Location/Qualifiers  
1..277  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A630094E02"  
/clone\_1lb="RIKEN full-length enriched, 3 days neonate  
thymus"  
/tissue-type="thymus"  
/dev\_stage="3 days neonate"



```

/lab.host="DHI08"
/notes="Site_1: Sali; Site_2: BanH; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research. In Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGAGGATCCAGAGCTCTTTTGTTTTGTTTTNN 3', cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence 15' GAGGAGAGAGATTCGAGTGAATTAATATCCCCCCCC
3'. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pluescript KS(+) after bulk excision from Lambda
FLC I."

```

	Query Match	3.7%; Score 75;	DB 10;	Length 277;
	Best Local Similarity	100.0%;	Pred. No. 8.4e-26;	
	Matches	75; Conservative	0; Mismatches	0; Indels
				Gaps 0;
OY	1851 GGAGACAGGTTTATTGGCAAAAGAATTA	CTGGTCGTGACAATGCACCAATATF	ACTG	1910
Dd	16 GGAACAGCAGGTTTTATTGGCAAAAGAA	TAACTGGTGTGACCAATGCACCAATAT	CTG	75
OY	1911 GCATATTCAGATAAC		1925	
Dd	76 GCATATTCAGATAAC		90	

RESULT 14	
BB215231	
LOCUS	312 bp. mRNA. linear
DEFINITION	BB215231 RIKEN full-length enriched, adult male aorta and vein Mus musculus cDNA A530025M07 3', mRNA sequence.
ACCESSION	BB215231

VERSION	BA215231.1	GI:8880184
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 312)

Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arahawa, T., Carninci

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iwama, M., Kadoori, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kuthirara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toyota, T., Tsunoda, Y., Wataniki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, T.  
RIKEN Mouse Estrs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
urk:http://genome.gsc.riken.go.jp/  
Carnilich,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki  
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Itawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, T. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```

FEATURES
SOURCE
Location/Vnat11iers
1. 312
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A530025M07"
/clone_1bp="RIKEN full-length enriched, adult male aorta
and vein"
/sex="male"
/tissue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Salt; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGGAGAGAGATCCAGAGAGCTCTTTTTTTTTTTTTTTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGGAGAGTTCTCAGATTAAATTATATATATATATATATATAT
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pbluescript KS(+) after bulk excision from lambda
FLC I."

```

BASE COUNT	84	a	88	c	56	g	84	t
ORIGIN								
Query Match			3.6%		Score 73;	DB 10;	Length 312;	
Best Local Similarity			100.0%		Pred. No. 9.9e-27;			
Matches 73;			Conservative 0;		Mismatches 0;		Indels 0;	Gaps 0;

Qy 480 CCCCTCCTTTACCTTCTGAATCCACAGTCAGAGAGCAAGTATTTCTCAACACAGA 539  
|||||  
104 CCCCTCCTTTACCTTCTGAATCCACAGTCAGAGAGCAAGTATTTCTCAACACAGA 163

QY	540	CAAGGAATGCTG	552
Db	164	CAAGGAATGCTG	176

RESULT 15	
AV160656	
LOCUS	
DEFINITION	
ACCESSION	
AV160656	246 bp mRNA
AV160656	Mus musculus head C57BL/6J 12-day embryo linear
clone 3010052H13, mRNA sequence.	EST 07-JUL-1999
AV160656	Mus musculus cDNA

VERSION	AV1606056.1	GI:5404291
KEYWORDS	EST.	
SOURCE	house mouse,	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus. 1 (bases 1 to 246)	
AUTHORS	Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayashi,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., NiiSUMA,H., Oda,H., Owa,C.,	



eye library CV1) corresponding to plates R-CV0-BRH through R-CV0-BRC, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTR, and R-CV1-BVU through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BV0P (7-9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0P-AOI through R-BV0P-AOK, and R-BX0-NOY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CY0, CZ0, DAO, DB0, DC0, DDO, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BXA through R-CZ0-BYL, R-CZ0-BZB-C, R-DAO-BX1 through R-DAO-BXP, R-DAO-BZD through R-DAO-BZH, R-DB0-BYQ through R-DB0-BZA, R-DC0-BZ1 through R-DC0-BZQ, R-DC0-CAY through R-DC0-CBA, R-DD0-BZB through R-DD0-BZC, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CNO, that makes up the tracer. The corresponding plates are R-CNO-BKW through R-CNO-BLD, R-CNO-BLG, R-CNO-BLP through R-CNO-BLR, R-CNO-BLT, R-CNO-BLM-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CNO pool corresponding to the following addresses: bkw-a-09-0-UI, bkw-b-09-0-UI, bkw-b-11-0-UI, bkw-b-10-0-UI, bxx-d-01-0-UI, bxx-d-06-0-UI, bxx-g-08-0-UI, bxx-h-12-0-UI, bky-a-05-0-UI, bkz-a-06-0-UI, bkz-a-11-0-UI, bkz-c-06-0-UI, bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blq-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CNO clone (corresponding to the address bkz-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population. TAG-lib-UI-R-CN1 TAG-TISSUE=fundus TAG\_SEQ=TTGCG" BASE COUNT 107 a 96 c 90 g 143 t

Query Match 2.6%; Score 53; DB 14; Length 436; Best Local Similarity 100.0%; Pred. No. 3,9e-16; Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1765 AACGAGCCATCAACCTCTCTGAGAAATGAGAGACTGTGAGC 1817  
 |||||||  
 Db 366 AACGAGCCATCAACCTCTCTGAGAAATGAGAGACTGTGAGC 314  
 |||||||

RESULT 17  
 BE531642 789 bp mRNA linear EST 09-AUG-2000  
 LOCUS 601230651f1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3594612 5',  
 DEFINITION mRNA sequence.  
 BE531642  
 ACCESSION BE531642.1 GI:9760287  
 VERSION  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 789)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbri@mail.nih.gov](mailto:cgapbri@mail.nih.gov)  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM8769 row: e column: 13  
 High quality sequence stop: 587.  
 Location/Qualifiers  
 1..789  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3594612"  
 /clone\_1lb="NCI\_CGAP\_Mam6"  
 /sex="female, virgin"  
 /tissue.type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-Sport6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 235 a 180 c 225 g 149 t

ORIGIN

Query Match 2.4%; Score 50; DB 10; Length 789; Best Local Similarity 100.0%; Pred. No. 1.8e-14; Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 852 CATGTGATGGCCAGCGAGCTGACAGACATGTTGAGACACACCA 901  
 |||||||  
 Db 515 CATGTGATGGCCAGCGAGCTGACAGACATGTTGAGACACACCA 564  
 |||||||

RESULT 18  
 BE209882 335 bp mRNA linear EST 21-APR-2001  
 LOCUS RST29410 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE209882  
 ACCESSION BE209882  
 VERSION BE209882.1 GI:13731569  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 335)  
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramchandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J. and Ducar,M.  
 TITLE Creation of genome-wide protein expression libraries using random activation of gene expression  
 Nat. Biotechnol. 19 (5), 440-445 (2001)  
 JOURNAL MEDLINE  
 COMMENT 21227151  
 CONTACT: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9536  
 Email: [scaine@atersys.com](mailto:scaine@atersys.com)  
 High quality sequence stop: 285.  
 Location/Qualifiers  
 1..335  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1lb="Athersys RAGE Library"  
 /cell\_line="HT1080"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation method was used, these sequence tags are not necessarily expressed in H1080 under normal circumstances."

BASE COUNT 95 a 63 c 74 g 101 t 2 others

ORIGIN

Query Match 2.3%; Score 47; DB 12; Length 335;  
Best Local Similarity 100.0%; Pred. No. 5.1e-13;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGAATCTACAGATTAAGATTTTGGACACATCGAAGCAGA 1979  
|||||  
66 CCTGAATCTACAGATTAAGATTTTGGACACATCGAAGCAGA 112

Db

RESULT 19 436 bp mRNA linear EST 19-NOV-2001  
BB817265 RIKEN full-length enriched, lung RCB-0558 LLC CDNA Mus  
BB817265 musculus cDNA clone G730038N06 3', mRNA sequence.  
BB817265  
BB817265.1 GI:16989894

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 436)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 436)  
Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watanahi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.  
Location/Qualifiers  
1. 436

FEATURES  
source

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G730038N06"  
/clone\_lib="RIKEN full-length enriched, lung RCB-0558 LLC  
cDNA"  
/tissue\_type="lung"  
/cell\_line="RCB-0558 LLC"  
/note="pooled cell lines : (cell\_line=RCB-1751 WEHI 164),  
(cell\_line=CRU-216 JC), (cell\_line=RCB-0035 WEHI-3),  
(cell\_line=RCB-0464 Meth-A), (cell\_line=RCB-0545 OHTA),  
(cell\_line=RCB-0559 K-1. F1), (cell\_line=RCB-1283 B16  
melanoma), (cell\_type=B cells, cell\_line=CRU-1703 WEHI 231  
), (cell\_type=Leidy cells, cell\_line=CRU-2065 MRC-1),  
(cell\_type=Nullipotent stem cell, cell\_line=CRU-2070 NE),  
(tissue\_type=bladder, cell\_line=RCB-0544 MBT-2),  
(tissue\_type=bone marrow, cell\_type=stroma cell,  
cell\_line=CRU-2028 SR-4987), (tissue\_type=colon,  
cell\_line=RCB-0549 Cle-H3), (tissue\_type=kidney,  
cell\_line=CCU-142 RAG), (tissue\_type=submandibular gland,  
cell\_line=CRU-1734 SCA-9 clone 15), (strain=BALB/C,  
cell\_type=B cells, cell\_line=CRU-1669 BCL1 clone 13.20-3B3  
), (strain=C3H, tissue\_type=brain, cell\_line=CRU-1443  
BC3H1)"

BASE COUNT 131 a 116 c 72 g 117 t

ORIGIN

Query Match 2.3%; Score 47; DB 10; Length 436;  
Best Local Similarity 100.0%; Pred. No. 5.6e-13;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 300 CCCAGCAGGGTTTTCATATACACTATATACATCGCAGCAGG 346  
|||||  
1 CCCAGCAGGGTTTTCATATACACTATATACATCGCAGCAGG 47

Db

RESULT 20 484 bp mRNA linear EST 21-FEB-2001  
BG289059 RIKEN full-length enriched, lung RCB-0558 LLC  
BG289059  
BG289059.1 GI:13044522

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 484)  
NIH-MGC http://mgc.ncl.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLAM10398 row: b column: 09  
High quality sequence stop: 484.  
Location/Qualifiers  
1. 484

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4512680"  
/clone\_lib="NIH-MGC\_93"  
/tissue\_type="transitional cell papilloma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: bladder; Vector: PCMV-SPORE; Site: 1; Note:  
Site 2: Salt; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.7 kb. Library enriched for

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

NIH-MGC  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: LLAM10398 row: b column: 09  
High quality sequence stop: 484.  
Location/Qualifiers  
1. 484

FEATURES  
source



Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>

Plate: LLM1167 row: e column: 20  
High quality sequence start: 3  
High quality sequence stop: 718.  
Location/Qualifiers

## FEATURES

source

1. 722

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="5264635"

/clone\_lib="NIH\_MGC\_95"

/tissue\_type="hippocampus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag

); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',

size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT

214 a 154 c 168 g 186 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 47; DB 13; Length 722;  
Pred. No. 6.8e-13;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1933 CCTGGAATCTACAGATATGATATTTGGACACATCGAAGCAGCA 1979

Db

516 CCGGAACTCTACAGATATGATATTTGGACACATCGAAGCAGCA 562

RESULT 24

BE387191

LOCUS

601277054F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3618180 5',

DEFINITION

RNA sequence.

ACCESSION

BE387191

VERSION

BE387191.1 GI:9332556

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 813)

TITLE

NIH-MGC <http://mgc.nci.nih.gov/>.

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)

CONTACT

Contact: Robert Strausberg, Ph.D.

EMAIL

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement

ATCC/DC/DMP

CDNA Library Preparation

Ling Hong/Rubin Laboratory

DNA Sequencing

by: Incyte Genomics, Inc.

Clone distribution

MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at: [image.lnl.gov](http://image.lnl.gov)

Plate: LLM286 row: k column: 13

High quality sequence start: 70

High quality sequence stop: 706.

Location/Qualifiers

1. 813

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="3618180"

/clone\_lib="NIH\_MGC\_20"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)."

/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dt priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAAGGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

230 a 155 c 205 g 223 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 47; DB 10; Length 813;  
Pred. No. 7.1e-13;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1933 CCTGGAATCTACAGATATGATATTTGGACACATCGAAGCAGCA 1979

Db

409 CCGGAACTCTACAGATATGATATTTGGACACATCGAAGCAGCA 455

RESULT 25

BG702763

LOCUS

602664640F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4817504 5',

DEFINITION

mRNA sequence.

ACCESSION

BG702763

VERSION

BG702763.1 GI:13974426

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 534)

TITLE

NIH-MGC <http://mgc.nci.nih.gov/>.

JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Unpublished (1999)

CONTACT

Contact: Robert Strausberg, Ph.D.

EMAIL

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement

Miklos Palkovics, M.D., Ph.D.

CDNA Library Preparation

Michael J. Brownstein (NHGRI), Shitaki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed

by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing

by: Incyte Genomics, Inc.

Clone distribution

MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: LLM10717 row: o column: 09

High quality sequence stop: 534.

Location/Qualifiers

1. 534

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="4817504"

/clone\_lib="NIH\_MGC\_95"

/tissue\_type="hippocampus"

/lab\_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified

pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag

); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',

size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched

for full-length clones and constructed using the

Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT

154 a 121 c 132 g 127 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 31; DB 12; Length 534;  
Pred. No. 0.00016;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1682 TGCAGCCAGCAAACTGGAATACAGAGTGGG 1712

Db 265 TGCAGCCAGCAAAACCTGATATACAGATGGG 295  
 |||||||||||||||||||||||||||||||

RESULT 26  
 LOCUS BE971187 806 bp mRNA linear EST 04-OCT-2000  
 DEFINITION 601651171P1 NIH\_MGC\_81 Homo sapiens cDNA clone IMAGE:3934649 5',  
 mRNA sequence.  
 BE971187  
 ACCESSION BE971187.1 GI:10584523  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 806)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: CLONETECH Laboratories, Inc.  
 cDNA Library Preparation: CLONETECH Laboratories, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L10M775 row: m column: 18  
 High quality sequence stop: 426.  
 Location/Qualifiers  
 1..806  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3934649"  
 /clone\_lib="NIH\_MGC\_81"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Organ: muscle (skeletal); Vector: pDNR-LTB  
 (Clontech); Site\_1: Sfil (ggcgccctggcc); Site\_2: Sfil  
 (ggccatataggc); 5' and 3' adaptors were used in cloning  
 as follows: 5' adaptor sequence: 5'-CAGCGCATTTATGGCC-3'  
 and 3' adaptor sequence:  
 5'-ATTCTAGAGCGGCGGCGGCGCATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size  
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained  
 inserts by PCR. This library was enriched for full-length  
 clones and was constructed by Clontech Laboratories (Palo  
 Alto, CA)."

BASE COUNT 222 a 216 c 187 g 181 t

ORIGIN

Query Match 1.5%; Score 31; DB 12; Length 806;  
 Best Local Similarity 100.0%; Pred. No. 0.00019;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1682 TGCAGCCAGCAAAACCTGATATACAGATGGG 1712  
 |||||||||||||||||||||||||||||||

Db 245 TGCAGCCAGCAAAACCTGATATACAGATGGG 275

RESULT 27  
 LOCUS B1560354 936 bp mRNA linear EST 05-SEP-2001  
 DEFINITION 603253902P1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5296080 5',  
 mRNA sequence.  
 B1560354  
 ACCESSION B1560354.1 GI:15447668  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 936)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shuraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L10M1749 row: d column: 01  
 High quality sequence stop: 666.  
 Location/Qualifiers  
 1..936  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="5296080"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescript (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 220 a 252 c 247 g 217 t

ORIGIN

Query Match 1.5%; Score 31; DB 13; Length 936;  
 Best Local Similarity 100.0%; Pred. No. 0.0002;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1682 TGCAGCCAGCAAAACCTGATATACAGATGGG 1712  
 |||||||||||||||||||||||||||||||

Db 524 TGCAGCCAGCAAAACCTGATATACAGATGGG 554

RESULT 28  
 LOCUS A1229932 503 bp mRNA linear EST 20-JAN-1999  
 DEFINITION EST226627 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone  
 REMC053 3' end, mRNA sequence.  
 A1229932  
 ACCESSION A1229932  
 VERSION A1229932.1 GI:3813819  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 503)  
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,  
 Kerlavage,A.R. and Adams,M.D.  
 Rat Genome Project: Generation of a Rat EST (RST) catalog & Rat  
 Gene Index  
 Unpublished (1998)  
 CONTACT: Lee, NH  
 The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@igr.org  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1..503  
 /organism="Rattus sp."

FEATURES  
 source

```

/db_xref="ATCC (Inhost):2037163"
/db_xref="taxon:10118"
/clone="REMCO53"
/clone_1ib="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/notes="vector: pT733Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      152 a      114 c      97 g      140 t
ORIGIN

Query Match      1.4%; Score 29; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1300 TTGGCCATAGCTGCTATCCCTGGGAATT 1328
          |||||||
DB      406 TTGGCCATAGCTGCTATCCCTGGGAATT 378

RESULT 29
BI553673      742 bp      mRNA      linear      EST 05-SEP-2001
LOCUS      603190594F1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:5262094 5',
DEFINITION      mRNA sequence.
ACCESSION      BI553673
VERSION      BI553673.1 GI:15440985
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 742)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@b-riemail.nih.gov
Tissue procurement: Miklos Palkovits, M.D., Ph.D.
CDNA library preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshitaki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11660 row: k column: 23
High quality sequence stop: 734.
Location/Qualifiers
1..742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5262094"
/clone_1ib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC library."
BASE COUNT      211 a      164 c      172 g      195 t
ORIGIN

Query Match      1.2%; Score 24; DB 13; Length 742;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1949 TAAGATATTTGGACACATCGCA 1972
          |||||||
DB      554 TAAGATATTTGGACACATCGCA 577

```

```

RESULT 30
AA373225      384 bp      mRNA      linear      EST 21-APR-1997
LOCUS      EST852231 HSC172 cells I Homo sapiens CDNA 5' end, mRNA sequence.
DEFINITION      AA373225
ACCESSION      AA373225
VERSION      AA373225.1 GI:2025545
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 384)
Adams,M.D., Kervilange,A.R., Fleischmann,R.D., Fulcher,R.A., Bult
C.J., Lee,N.H., Kitzness,E.F., Weinstein,K.G., Gocayne,J.D., White
O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Mel,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodex,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrite,A., Fischer,C., Hastings,G.A., He,M.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Weli,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of CDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Contact: Kervilange, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkelv@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1..384
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):177586"
/db_xref="taxon:9606"
/clone_1ib="HSC172 cells I"
/cell_type="fibroblast"
/dev_stage="fetal"
/notes="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI"
BASE COUNT      101 a      81 c      97 g      101 t      4 others
ORIGIN

Query Match      1.1%; Score 23; DB 9; Length 384;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1780 ACCCTCCCTCAGCTGTGAGAAATA 1802
          |||||||
DB      195 ACCCTCCCTCAGCTGTGAGAAATA 217

RESULT 31
A0976049      581 bp      DNA      linear      GSS 28-JAN-2000
LOCUS      A0976049/c
DEFINITION      RPCI-23-333M12.TV RPCI-23 Mus musculus genomic clone RPCI-23-333M12

```



```

ACCESSION   DNA sequence.
VERSION     AO976049
KEYWORDS    AO976049.1 GI:6806506
SOURCE      GSS.
ORGANISM    house mouse.
            Mus musculus
REFERENCE   1 (bases 1 to 581)
AUTHORS     Zhao,S., Nieman,W., Feldblum,T., Malek,J., Shatsman,S., Akintet
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroll,M., de Jong,P.
            and Fraser,C.M.
TITLE       Mouse BAC End Sequences from Library RPCI-23
JOURNAL     Unpublished (1999)
COMMENT     Other-GSSs: RPCI-23-333M12.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhaod@ligr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@edj.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (http://www.ligr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html)
            Plate: 333 row: M column: 12
            Seq primer: T7
            Class: BAC ends.

FEATURES             Source          Location/Qualifiers
     source            1..581
     /organism="Mus musculus"
     /strain="C57BL/6J"
     /db_xref="taxon:10090"
     /clone="RPCI-23-333M12"
     /clone_lib="RPCI-23"
     /sex="Female"
     /lab_host="DH10B"
     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
     brain genomic DNA was isolated and partially digested
     with a combination of EcoRI and EcoRI methylase. Size
     selected DNA was cloned into the pBACe3.6 vector at the
     EcoRI sites. The ligation products were transformed into
     DH10B electrocompetent cells (BRL Life Technologies)."
     BASE COUNT      160 a      129 c      130 g      162 t
     ORIGIN
Query Match      1.1%, Score 23; DB 17; Length 581;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1710 GCGAGAGTGGTTGAGCTTATAT 1732
      |||||||
Db      24 GCGAGAGTGGTTGAGCTTATAT 2
      |||||||

RESULT 32
DEFINITION      A1365378      450 bp      mRNA
LOCUS           A1365378
ACCESSION       A1365378
VERSION         A1365378.1 GI:4125067
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE	1 (bases 1 to 450)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs.r@email.nih.gov">cgapbs.r@email.nih.gov</a> Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E Consortium/LNLN at: <a href="http://www.bio.llnl.gov/bdnp/image/image.html">www.bio.llnl.gov/bdnp/image/image.html</a> Insert Length: 601 Std Error: 0.00 Seq primer: -40UP from GlDco High quality sequence stop: 364.
FEATURES	Location/Qualifiers
SOURCE	1..450 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2020876" /clone.lib="NCI CGAP CL11" /tissue_type="B-cell, chronic lymphocytic leukemia" /lab_host="DH10B" /note="Vector: p7R73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTCGAGCGGCAGCATGTGGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7R73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT	78 a 158 c 150 g 63 t 1 others
ORIGIN	
Query Match	1.1%; Score 22; DB 9; Length 450;
Best Local Similarity	100.0%; Pred. No. 8.5;
Matches	22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DY	922 CAGAAGCCCAAGACGCTGTATG 943
Db	317 CAGAAGGCCAAGACGCTGTATG 338       
RESULT 33	
LOCUS	AL692789/c 1501 bp mRNA linear EST 21-MAR-2002
DEFINITION	AL692789 NAB1 Anopheles gambiae CDNA clone NAB1-P03-H-03-5, mRNA sequence.
ACCESSION	AL692789
VERSION	AL692789.1 GI:19612674
KEYWORDS	EST.
SOURCE	African malaria mosquito.
ORGANISM	Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
REFERENCE	1 (bases 1 to 1501) Christophides,G.K., Blass,K., Zdobnov,E.M., Carmouche,R., Benes,V. and Kafatos,F.C. Anopheles gambiae EST, European Molecular Biology Laboratory Unpublished (2002)
TITLE	Contact: Christophides GK
JOURNAL	Fotis C. Kafatos laboratory
COMMENT	European Molecular Biology Laboratory Meyerhofstrasse 1, 69117 Heidelberg, Germany Tel.: +49 6221 387-440 Fax: +49 6221 387-306 Email: christopheemb1-heidelberg.de Plate: P03 row: H column: 03.

FEATURES  
source  
Location/Qualifiers  
1.1501  
/organism="Anopheles gambiae"  
/db\_xref="taxon:7165"  
/clone="NAP1-P03-H-03-5"  
/clone\_lib="NAP1"  
/lab\_host="E.coli DH10B"  
/note="Vector: pT73D-Pac (Pharmacia); Site\_1: NotI; Site\_2: EcoRI; ESTs sequenced from the T7 priming site that reads from the 5' end of cDNA. The NAP1 is a directionally cloned and normalized, oligo-T primed cDNA library constructed from a mixture of Anopheles gambiae developmental stages according to: Ronald, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery, Genome Research 6, 791-806."

BASE COUNT 341 a 379 c 352 g 361 t 68 others

ORIGIN

Query Match 1.0%; Score 21; DB 9; Length 1501;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 727 TTGTAGCAGGCTTGCTTCA 747  
Db 399 TTGTAGCAGGCTTGCTTCA 379

RESULT 34  
AZ656375 220 bp DNA linear GSS 14-DEC-2000  
LOCUS 1M0531F19R Mouse 10kb plasmid UUGCIM library Mus musculus genomic  
DEFINITION clone UUGCIM0531F19 R. DNA sequence.  
ACCESSION AZ656375  
VERSION AZ656375.1 GI:11793521  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 220)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0531 Row: F Column: 19  
Seq primer: CACACAGAAACAGCATGACAC  
Class: Plasmid ends  
High quality sequence stop: 220.  
Location/Qualifiers  
1.220  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGCIM0531F19"  
/clone\_lib="Mouse 10kb plasmid UUGCIM library"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: pMD22nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (914732114191AF129072-1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 66 a 46 c 57 g 51 t

ORIGIN

Query Match 1.0%; Score 20; DB 17; Length 220;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 937 CTGTATCCCTCTGCTCCCA 956  
Db 209 CTGTATCCCTCTGCTCCCA 190

RESULT 35  
AV210530 264 bp mRNA linear EST 30-OCT-1999  
LOCUS AV210530 RIKEN full-length enriched, adult male testis Mus musculus  
DEFINITION cDNA clone 1700124p10 3', mRNA sequence.  
ACCESSION AV210530  
VERSION AV210530.1 GI:6151067  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 264)  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al. 1999)  
Unpublished (1999)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki, N., Izawa, M., Watanabe, M., Okazaki, Y. and Hayashizaki, Y.  
Matsuyama, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.





RESULT 39  
A0696374/c 540 bp DNA linear GSS 06-JUL-1999  
LOCUS  
DEFINITION HS-2171.A2\_H01.T7C CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate-2171 Col=2 Row=0, DNA sequence.  
ACCESSION A0696374  
VERSION A0696374.1 GI:5386622  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 540)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 2171 row: 0 column: 2  
Seq primer: T7  
Class: BAC.ends  
High quality sequence stop: 540.  
Location/Qualifiers  
1. 540  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate-2171 Col=2 Row=0"  
/clone\_1lb="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in  
E-Coli DH10B"  
BASE COUNT 194 a 130 c 73 g 133 t 10 others  
ORIGIN

Query Match 1.0%; Score 20; DB 17; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1269 TGTGATGTCAGATGTTA 1288  
Db 417 TGTGATGTCAGATGTTA 398

RESULT 40  
A2794922 560 bp DNA linear GSS 16-FEB-2001  
LOCUS  
DEFINITION 2M0048119R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0048119 R. DNA sequence.  
ACCESSION A2794922  
VERSION A2794922.1 GI:12941393  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 560)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duvul,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL Plasmid Inserts  
COMMENT Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: rduffy@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0048 row: 1 column: 19  
Seq primer: CACGACGAAACAGCATGACC  
Class: plasmid ends  
High quality sequence stop: 560.  
Location/Qualifiers  
1. 560  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0048119"  
/clone\_1lb="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (g1147321149b1AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 151 a 115 c 184 g 110 t  
ORIGIN

Query Match 1.0%; Score 20; DB 17; Length 560;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 GGATTGCACGACACTGT 52  
Db 321 GGATTGCACGACACTGT 340

RESULT 41  
BB612587 648 bp mRNA linear EST 31-AUG-2001  
LOCUS  
DEFINITION BB612587 RIKEN full-length enriched, 0 day neonate skin Mus  
musculus cDNA clone 4632427H21 5', mRNA sequence.  
ACCESSION BB612587  
VERSION BB612587.1 GI:15395169  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 648)  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koyama,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,



REFERENCE 1 (bases 1 to 822)  
 AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shwartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P. and Fraser, C. M.  
 TITLE Mouse BAC End Sequences from Library RPCI-24  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: RPCI-24-317N21.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhaoc@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.tigr.org/bacpac/orderingframe.html>). BAC end page: [http://www.tigr.org/db/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/db/bac_ends/mouse/bac_end_intro.html)  
 Plate: 317 row: N column: 21  
 Seq primer: SP6  
 Class: BAC ends.  
 FEATURES  
 Location/Qualifiers  
 1..822  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-317N21"  
 /clone\_1lb="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."  
 BASE COUNT 184 a 182 c 209 g 247 t  
 ORIGIN  
 Query Match 1.0%; Score 20; DB 17; Length 822;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 502 CCACAGTCAGAGAGCAAG 521  
 ||||||||||||||||||  
 DB 154 CCACAGTCAGAGAGCAAG 135  
 RESULT 44  
 CNS07928 1018 bp DNA linear GSS 08-JUL-2001  
 LOCUS T7 end of clone XBB0AA002C11 of library XBB0AA from strain CBS 4732  
 DEFINITION of *Pichia angusta*, genomic survey sequence.  
 ACCESSION AL435930  
 VERSION AL435930.1 GI:12219343  
 KEYWORDS GSS.  
 SOURCE *Pichia angusta*.  
 ORGANISM *Pichia angusta*.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; *Pichia*.  
 REFERENCE 1 (bases 1 to 1018)  
 AUTHORS Souciet, J. L., Aigle, M., Artiguenave, F., Blandin, G., Boloitin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 1018)

AUTHORS Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*  
 JOURNAL FEBS Lett. 487 (1), 76-81 (2000)  
 MEDLINE 20584723  
 PUBMED 11152868  
 REFERENCE 3 (bases 1 to 1018)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: [sequef@genoscope.cns.fr](mailto:sequef@genoscope.cns.fr) - Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
 FEATURES  
 Location/Qualifiers  
 1..1018  
 /organism="Pichia angusta"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4805"  
 /clone="XBB0AA002C11"  
 /clone\_1lb="XBB0AA"  
 /note="end: 17"  
 /note="complement(17...>717)"  
 /note="similar to Saccharomyces cerevisiae ORF YK1085w (MDH1; maleate dehydrogenase precursor, mitochondrial)"  
 /note="1 putative frameshift(s)"  
 /evidence="not\_experimental"  
 BASE COUNT 241 a 239 c 315 g 219 t 4 others  
 ORIGIN  
 Query Match 1.0%; Score 20; DB 17; Length 1018;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1535 TCAGAGACCTGCTAAGCA 1554  
 ||||||||||||||||||  
 DB 375 TCAGAGACCTGCTAAGCA 356  
 Search completed: July 5, 2003, 15:48:41  
 Job time : 2982 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 10:00:34 ; Search time 467 Seconds

(without alignments)  
9880.823 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttccagtgcctactacttg.....ttgaagtgcctactacttg 2049

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 20

Total number of hits satisfying chosen parameters: 25

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	100.0	2049	21	AAA97634	Mouse mature neutr
2	2049	100.0	2271	21	AAA97633	Mouse neutral/alka
3	2049	100.0	3108	21	AAA97632	Mouse neutral/alka
4	2016	98.4	4835	21	AAA97635	Mouse neutral/alka
5	329	16.1	551	22	AAH98033	Murine 7-transmemb
6	68	3.3	2283	22	AAF99977	DNA encoding rat p
7	65	3.2	65	24	ABN53673	Mouse spliced tran
8	53	2.6	539	22	AAF99979	PCR product isolat
9	47	2.3	2507	24	AAD22668	Human ceramidase c

10	43	2.1	3058	24	ABL99947	Human slaly1 trans
11	38	1.9	38	21	AAA97627	Mouse neutral/alka
12	31	1.5	299	21	AAO03574	Human secreted pro
13	29	1.4	1194	20	AAZ06239	Human secreted pro
14	25	1.2	38	24	AAD22672	Human ceramidase c
15	24	1.2	24	21	AAA97636	Mouse neutral/alka
16	24	1.2	24	21	AAA97637	Mouse neutral/alka
17	22	1.1	4020	19	AAV41320	Human angiotensin
18	22	1.1	4020	21	AAA38330	Human angiotensin
19	22	1.1	4022	24	AAK99395	DNA of APP related
20	22	1.1	4024	11	AAQ04027	Human angiotensin
21	22	1.1	4024	22	AAK35850	Human angiotensin
22	22	1.1	5005	22	AAH57430	Human intestine ce
23	21	1.0	21	22	AAF99978	PCR primer used to
24	21	1.0	3942	20	AAK35851	Rat angiotensin co
25	20	1.0	5681	22	AAK68974	Human immune/haema

#### ALIGNMENTS

RESULT 1  
ID AAA97634 standard; DNA; 2049 BP.  
AC AAA97634;  
DT 02-FEB-2001 (first entry)  
DE Mouse mature neutral/alkaline ceramidase-encoding DNA, SEQ ID NO:15.  
KW Neutral/alkaline ceramidase; mouse; murine; recombinant production;  
KW cellular ceramide content control; antibody; membrane synthesis;  
KW lipid engineering; ceramide metabolism; drug development;  
KW mature ceramidase; ds.  
OS Mus sp.  
PN W0200058448-A1.  
PD 05-OCT-2000.  
PF 24-MAR-2000; 2000WO-JP01802.  
PR 26-MAR-1999; 99JP-0084743.  
PA (TAKI ) TAKARA SHUZO CO LTD.  
PI Ito M;  
PS WPI: 2000-619079/59.  
P-PSDB: AAB23292.  
PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering reagent for studying structure and functions of ceramide as well as developing drugs for diseases associated with ceramide metabolism -  
PS Claim 1: Page 65-66; 76pp; Japanese.  
CC The invention relates to a mouse neutral/alkaline ceramidase and to nucleic acids encoding it. The invention also relates to expression vectors and host cells comprising DNA encoding mouse neutral/alkaline ceramidase, the recombinant production of the ceramidase, an antibody against the ceramidase, and a method for controlling the ceramide content in cells and/or tissues by introducing the gene or its antisense nucleic acid into the cells and/or tissues. The ceramidase can be used as a lipid engineering reagent for studying structure and functions of ceramide as well as developing drugs for diseases associated with ceramide metabolism. The present sequence represents DNA encoding the mature mouse neutral/alkaline ceramidase.  
CC Sequence 2049 BP; 589 A; 470 C; 490 G; 500 T; 0 other;  
SQ

Query Match 100.0%; Score 2049; DB 21; Length 2049;  
Best Local Similarity 100.0%; Pred.No. 0;  
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCAGTGGCTACTACATTGGCGTTGGAGAGCGGATTTGCACAGACAGTGCATGATTC 60  
DB 1 TTCAGTGGCTACTACATTGGCGTTGGAGAGCGGATTTGCACAGACAGTGCATGATTC 60  
QY 61 AATTGTAGTGGCTATGGCAAAAATGGCCAGAAATGCACGGGGTCTCTCACAGCGCTGTT 120  
DB 61 AATTGTAGTGGCTATGGCAAAAATGGCCAGAAATGCACGGGGTCTCTCACAGCGCTGTT 120  
QY 121 ACCCGTCTTTATATCTGGCGGATCCAGATGGGTCAATTCGAATGGCATTTGTGACGCTG 180  
DB 121 ACCCGTCTTTATATCTGGCGGATCCAGATGGGTCAATTCGAATGGCATTTGTGACGCTG 180  
QY 181 GAATATGTATGATTTCCCAACGACTGAGTGGAGGCTCTGAAGAGCTATAGAGATAA 240  
DB 181 GAATATGTATGATTTCCCAACGACTGAGTGGAGGCTCTGAAGAGCTATAGAGATAA 240  
QY 241 TATGGCTCTCTGTATCGAAGACAAATGTTATCTGAGTGCATTCACACACACTGTGC 300  
DB 241 TATGGCTCTCTGTATCGAAGACAAATGTTATCTGAGTGCATTCACACACACTGTGC 300  
QY 301 CCAGAGAGGTTTTCGAATATACACTCTATATCTGCGCAGAGGAGATTCAGCAACCGG 360  
DB 301 CCAGAGAGGTTTTCGAATATACACTCTATATCTGCGCAGAGGAGATTCAGCAACCGG 360  
QY 361 ACCTTACAGTACATAGTCTCTGGATCATGAAGAGCATTTAGCTACACAAATCTT 420  
DB 361 ACCTTACAGTACATAGTCTCTGGATCATGAAGAGCATTTAGCTACACAAATCTT 420  
QY 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGCAGATCAACGAGAC 480  
DB 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGCAGATCAACGAGAC 480  
QY 481 CCCCTCTCTTACTCTGTGATCGACAGTCAGAGAGAGCAAGATATCTTCAAAACAGAC 540  
DB 481 CCCCTCTCTTACTCTGTGATCGACAGTCAGAGAGAGCAAGATATCTTCAAAACAGAC 540  
QY 541 AAGGAATATGCTGCTTGAACCTGTGATTTGAATGAGAGACTTGGGTCTTATCACG 600  
DB 541 AAGGAATATGCTGCTTGAACCTGTGATTTGAATGAGAGACTTGGGTCTTATCACG 600  
QY 601 TGGTTTGGCATTCACCCCGTGAAGTGAACATAGCAACCTTGTGATAGTGCAAT 660  
DB 601 TGGTTTGGCATTCACCCCGTGAAGTGAACATAGCAACCTTGTGATAGTGCAAT 660  
QY 661 ATGGGCTATGGGCTTACCTTTTGAAGCAAAAGAAACAAAGGCTATGCTGGACAG 720  
DB 661 ATGGGCTATGGGCTTACCTTTTGAAGCAAAAGAAACAAAGGCTATGCTGGACAG 720  
QY 721 GGACGCTTGTAGACAGGCTTGTCTCATCAAAATCTCGAGAGCTGTCAACCAATCTT 780  
DB 721 GGACGCTTGTGTAGACAGGCTTGTCTCATCAAAATCTCGAGAGCTGTCAACCAATCTT 780  
QY 781 GGGCCGCAATTTGTGTACACAGAGGAGTCTTGTGACACGACAGACACCTGTCCCAAC 840  
DB 781 GGGCCGCAATTTGTGTACACAGAGGAGTCTTGTGACACGACAGACACCTGTCCCAAC 840  
QY 841 GGTGGGCTATGAGATGGCATGGCGAGGAGCTGGCAAGAGATGTTTGAAGGACACAG 900  
DB 841 GGTGGGCTATGAGATGGCATGGCGAGGAGCTGGCAAGAGATGTTTGAAGGACACAG 900  
QY 901 AATTATAGACGAGATCATATCAGAAAGGCAAGAGAGCTGTATGCTCTGCCAGAG 960  
DB 901 AATTATAGACGAGATCATATCAGAAAGGCAAGAGAGCTGTATGCTCTGCCAGAG 960  
QY 961 GTGACCGGCGCAGTCTTGTACACAGTGGGTGAACATGACAGATGTGAGCGCTCAG 1020  
DB 961 GTGACCGGCGCAGTCTTGTACACAGTGGGTGAACATGACAGATGTGAGCGCTCAG 1020  
QY 1021 CTCATATGCCACACACAGATGAAGACGTCTAAACCTGCCCTGGCTACAGTTTGGCGCA 1080

DB 1021 CTCATATGCCACACACAGATGAAGACGTCTAAACCTGCCCTGGCTACAGTTTGGCGCA 1080  
QY 1081 GGCACAATTGATGAGATTTGGGCGCTCAATATTTACACAGGAACTACGGAAGGGATCCA 1140  
DB 1081 GGCACAATTGATGAGATTTGGGCGCTCAATATTTACACAGGAACTACGGAAGGGATCCA 1140  
QY 1141 TTCTGGACACTCTTCGGGACACAGCTCTTGGGAAAACATCTGAAGAGATTTAGAGTGT 1200  
DB 1141 TTCTGGACACTCTTCGGGACACAGCTCTTGGGAAAACATCTGAAGAGATTTAGAGTGT 1200  
QY 1201 CAGAAACCAACCAATCTGCTTCAACATGAGAGCTGACATACCATCTGGGCA 1260  
DB 1201 CAGAAACCAACCAATCTGCTTCAACATGAGAGCTGACATACCATCTGGGCA 1260  
QY 1261 CCAGATATTTGATGATTTGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1320  
DB 1261 CCAGATATTTGATGATTTGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1320  
QY 1321 GGGGAATTAACAACCATGTCGGGACGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380  
DB 1321 GGGGAATTTAACAACCATGTCGGGACGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380  
QY 1381 CTTTATGGATGAAGATATGACCTGTTATCGAGCTTAAGCAATGTTTATACAT 1440  
DB 1381 CTTTATGGATGAAGATATGACCTGTTATCGAGCTTAAGCAATGTTTATACAT 1440  
QY 1441 TACATTTACACATATGAAATATACAGGCTGACGCTAGAGAGACATCTCAATCTAT 1500  
DB 1441 TACATTTACACATATGAAATATACAGGCTGACGCTAGAGAGACATCTCAATCTAT 1500  
QY 1501 GGCACACACACCTCTGCTGCATATCATCACTCTTACAGACCTTCTTAAGCAATGCT 1560  
DB 1501 GGCACACACACCTCTGCTGCATATCATCACTCTTACAGACCTTCTTAAGCAATGCT 1560  
QY 1561 ACGGACACAGTGGCCCAATGAGAGAGTGGTCCGAGCCCTCCATTTCTTCAAAATCTAAT 1620  
DB 1561 ACGGACACAGTGGCCCAATGAGAGAGTGGTCCGAGCCCTCCATTTCTTCAAAATCTAAT 1620  
QY 1621 GCTTCACTTATTCCTAATATTTGCGGATAGAGACCAATTTGGCAAACTTTGGGATGTC 1680  
DB 1621 GCTTCACTTATTCCTAATATTTGCGGATAGAGACCAATTTGGCAAACTTTGGGATGTC 1680  
QY 1681 TTGCAGCCGCAAAACCTTAATACAGAGTGGGAGAAAGTGTGAAGTTATATTTGTAGGC 1740  
DB 1681 TTGCAGCCGCAAAACCTTAATACAGAGTGGGAGAAAGTGTGAAGTTATATTTGTAGGC 1740  
QY 1741 GCTAACCCAAAGAAATTCACAGAGAAACAGACCCATCAAACTTCTCTCATCTGTGAGAA 1800  
DB 1741 GCTAACCCAAAGAAATTCACAGAGAAACAGACCCATCAAACTTCTCTCATCTGTGAGAA 1800  
QY 1801 TAGGAGACTCTGTAGCTGACTGGCAGATATGTATACGATGCTCTGGAGACGAGG 1860  
DB 1801 TAGGAGACTCTGTAGCTGACTGGCAGATATGTATACGATGCTCTGGAGACGAGG 1860  
QY 1861 TTTTATTTGACAAAGGATATCTGGGCTGTGACCAATTCGAACAATATCTGCAATTTCCA 1920  
DB 1861 TTTTATTTGACAAAGGATATCTGGGCTGTGACCAATTCGAACAATATCTGCAATTTCCA 1920  
QY 1921 GATACCTGCTTACCTGGAATTCAGAAATPAATATTTTGGACACAATGGAAGAGAA 1980  
DB 1921 GATACCTGCTTACCTGGAATTCAGAAATPAATATTTTGGACACAATGGAAGAGAA 1980  
QY 1981 CTTTGAACCCGCTGTCTATACATATTTGAAGAAATTTCTTCTCTTTGAAGTGTG 2040  
DB 1981 CTTTGAACCCGCTGTCTATACATATTTGAAGAAATTTCTTCTCTTTGAAGTGTG 2040  
QY 2041 ACTACTTAG 2049  
DB 2041 ACTACTTAG 2049

RESULT 2

AAA97633  
ID AAA97633 standard; DNA; 2271 BP.  
XX  
AC AAA97633;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Mouse neutral/alkaline ceramidase ORF. SEQ ID NO:12.  
XX  
KM Neutral/alkaline ceramidase; mouse; murine; recombinant production;  
KM cellular ceramide content control; antibody; membrane synthesis;  
KM lipid engineering; ceramide metabolism; drug development;  
KM open reading frame; ORF; ds.  
XX  
OS Mus sp.  
XX  
PN M0200058448-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 24-MAR-2000; 2000MO-JP01802.  
XX  
PR 26-MAR-1999; 99JP-0084743.  
XX  
PA (TAKI ) TAKARA SHUZO CO LTD.  
XX  
PI Ito M;  
XX  
DR WPI: 2000-619079/59.  
XX  
DR P-PSDB: AAB23291.  
XX  
PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
PT reagent for studying structure and functions of ceramide as well as  
PT developing drugs for diseases associated with ceramide metabolism -  
XX  
PS Example 4; Page 57-58; 76pp; Japanese.  
XX  
CC The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents the mouse  
CC neutral/alkaline ceramidase open reading frame (ORF).  
XX  
SQ Sequence 2271 BP; 647 A; 541 C; 525 G; 558 T; 0 other;

Query Match 100.0%; Score 2049; DB 21; Length 2271;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TTCAGTGGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGTACAGATATC 60  
|||||  
223 TTCAGTGGCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGTACAGATATC 282  
61 AATTGATGGGCTATGCGCAAAATGCGCAATGACAGGGGTCTCTCCACAGGCTGTTC 120  
|||||  
283 AATTGATGGGCTATGCGCAAAATGCGCAATGACAGGGGTCTCTCCACAGGCTGTTC 342  
121 AGCGTGGCTTATCTGGCGGATGCAGATGGGTCAAAATGCAATGGCATTTGTGACGCTG 180  
|||||  
343 AGCGTGGCTTATCTGGCGGATGCAGATGGGTCAAAATGCAATGGCATTTGTGACGCTG 402  
181 GAACATATGATGATTTCCCAACGACGTAGTGGAGGCTCTGAAGAGACTAGAGATAA 240  
|||||  
403 GAACATATGATGATTTCCCAACGACGTAGTGGAGGCTCTGAAGAGACTAGAGATAA 462  
241 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCATTTCACACACACTCTGGC 300  
|||||

463 TATGGCTCTCTATGCAAGACAAATGTTATCTGAGTGCCATTCCACACACTCTGGC 522  
DB  
301 CCAGCAGGGTTTTTCCAAATATACCTCTATATCTGCGAGGAGGATTCGCAACCGG 360  
QY  
523 CCAGCAGGGTTTTTCCAAATATACCTCTATATCTGCGAGGAGGATTCGCAACCGG 582  
DB  
361 ACCTTTCAGTACATAGTCTCTGGGATCATGAAGAGCATTTGATATAGTCCACAAATCTT 420  
QY  
583 ACCTTTCAGTACATAGTCTCTGGGATCATGAAGAGCATTTGATATAGTCCACAAATCTT 642  
DB  
421 AAACGAGGCAAAATCTTTATCAACAAGAAATGTTGCTAATGTGCATCAACCGAGC 480  
QY  
643 AAACGAGGCAAAATCTTTATCAACAAGAAATGTTGCTAATGTGCATCAACCGAGC 702  
DB  
481 CCCTCTCTTACCTCTGATCCACAGTCAGAGAGGCAAGGATTTCTCAACACAGAC 540  
QY  
703 CCCTCTCTTACCTCTGATCCACAGTCAGAGAGGCAAGGATTTCTCAACACAGAC 762  
DB  
541 AAGGAATGCTGCTTTGAACCTGATGATTTGAATGAGAGACTTGGGCTTTATCACG 600  
QY  
763 AAGGAATGCTGCTTTGAACCTGATGATTTGAATGAGAGACTTGGGCTTTATCACG 822  
DB  
601 TGGTTGGCATCCACCCCGTGAACATGAACAAATAGCACCTTGTGTAATGTACAAAT 660  
QY  
823 TGGTTGGCATCCACCCCGTGAACATGAACAAATAGCACCTTGTGTAATGTACAAAT 882  
DB  
661 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAACAAAGGCTATGCTGACAGC 720  
QY  
883 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAACAAAGGCTATGCTGACAGC 942  
DB  
721 GGACGCTTTGTAGCAGGCTTTGCTTCATCAAAATCTGAGAGCTGTACCCAAATCTT 780  
QY  
943 GGACGCTTTGTAGCAGGCTTTGCTTCATCAAAATCTGAGAGCTGTACCCAAATCTT 1002  
DB  
781 GGCCCGCATTTGTGTCAACACAGGGAGTTTGTGACACGACGACAGACCTGTCCCAAC 840  
QY  
1003 GGCCCGCATTTGTGTGTCAACACAGGGAGTTTGTGACACGACGACAGACCTGTCCCAAC 1062  
DB  
841 GGTTGGGCTTATGATGTCATGGCGGAGGAGCTGGCAAGACATGTTTGAAGACACAC 900  
QY  
1063 GGTTGGGCTTATGATGTCATGGCGGAGGAGCTGGCAAGACATGTTTGAAGACACAC 1122  
DB  
901 ATTATAGCAGGATCATCTATCAGAAAGGCAAGAGCTTATGCTCTGCTCCAGAG 960  
QY  
1123 ATTATAGCAGGATCATCTATCAGAAAGGCAAGAGCTTATGCTCTGCTCCAGAG 1182  
DB  
961 GTGACCGGCCAGTGTCTTACAGTCAACAGTGGTGTAAATGACAGATGTGACGCTCAG 1020  
QY  
1183 GTGACCGGCCAGTGTCTTACAGTCAACAGTGGTGTAAATGACAGATGTGACGCTCAG 1242  
DB  
1021 CTCGAATGCCACACACAGTGAAGAGGCTGTAAACGCTGGGCTACAGTTTGGCGCA 1080  
QY  
1243 CTCGAATGCCACACACAGTGAAGAGGCTGTAAACGCTGGGCTACAGTTTGGCGCA 1302  
DB  
1081 GGCACAATTTGATGAGTTCGGGCTCAATATTATACAGAGGAATACGGAAGGGATCCA 1140  
QY  
1303 GGCACAATTTGATGAGTTCGGGCTCAATATTATACAGAGGAATACGGAAGGGATCCA 1362  
DB  
1141 TTCTGGACACCTCTTGGGACACAGCTTTGGGAAACCATCTGAAAGAGATTGTAGAGT 1200  
QY  
1363 TTCTGGACACCTCTTGGGACACAGCTTTGGGAAACCATCTGAAAGAGATTGTAGAGT 1422  
DB  
1201 CAGAAACCCAAACCAATCTGCTTACAGATGGAGAGCTAGCAATACCATCTGGGCA 1260  
QY  
1423 CAGAAACCCAAACCAATCTGCTTACAGATGGAGAGCTAGCAATACCATCTGGGCA 1482  
DB  
1261 CCAGATATTGTTGATGTTGATGTTGTTTACGTTGGGCTCTTGGCATATGCTCTATCCCT 1320  
QY  
1483 CCAGATATTGTTGATGTTGATGTTGTTTACGTTGGGCTCTTGGCATATGCTCTATCCCT 1542  
DB  
1321 GGGGAATTAACAACCATGTGCGGACGAGAGATTTCTGTGAGCAATTAAGAAATTTGCA 1380  
QY  
1543 GGGGAATTAACAACCATGTGCGGACGAGAGATTTCTGTGAGCAATTAAGAAATTTGCA 1602  
DB

QY 1381 CTTATGAGATGAAGATATGACCGTTGTTATCGAGGCTCAAGCAATGTTATACACAT 1440  
Db 1603 CTTATGAGATGAAGATATGACCGTTGTTATCGAGGCTCAAGCAATGTTATACACAT 1662  
QY 1441 TACATTACCAATATGAAGATACCAGGCTCAGGGGTACGAGGACGATCTACAACTAT 1500  
Db 1663 TACATTACCAATATGAAGATACCAGGCTCAGGGGTACGAGGACGATCTACAACTAT 1722  
QY 1501 GGACACACACCCCTGTCTGATATTCACCTTTACAGAGCTTGTCAAGCAATTCCT 1560  
Db 1723 GGACACACACCCCTGTCTGATATTCACCTTTACAGAGCTTGTCAAGCAATTCCT 1782  
QY 1561 ACGGACACAGTAGCCAAACATGAGAGTGGTCCGAGCCCTCCATCTTCAAAAATCTAATA 1620  
Db 1783 ACGGACACAGTAGCCAAACATGAGAGTGGTCCGAGCCCTCCATCTTCAAAAATCTAATA 1842  
QY 1621 GCTTCACTTATCTCTAATATTCGCGATAGACACCAATTTGGCAAAATTTGGGATGTC 1680  
Db 1843 GCTTCACTTATCTCTAATATTCGCGATAGACACCAATTTGGCAAAATTTGGGATGTC 1902  
QY 1681 TTGACGACGAGCAAAACCTGAATACAGAGTGGGAGAGTGTGAAGTTATATTTGTAGC 1740  
Db 1903 TTGACGACGAGCAAAACCTGAATACAGAGTGGGAGAGTGTGAAGTTATATTTGTAGC 1962  
QY 1741 GCTAACCCAAAGATTTACGAGAGACAGACCCATCAACCTTCTACTGTGAGAA 1800  
Db 1963 GCTAACCCAAAGATTTACGAGAGACAGACCCATCAACCTTCTACTGTGAGAA 2022  
QY 1801 TACGAGAGCTGTAGCTGAGTGGAGATATATATATGATATGATGCTCTGGGAGAGAGG 1860  
Db 2023 TACGAGAGCTGTAGCTGAGTGGAGATATATATATGATATGATGCTCTGGGAGAGAGG 2082  
QY 1861 TTTTATTTGGCAAAAGAACTACTGAGTGGTCTGAGCAATGCAACATATATCTGCATATTC 1920  
Db 2083 TTTTATTTGGCAAAAGAACTACTGAGTGGTCTGAGCAATGCAACATATATCTGCATATTC 2142  
QY 1921 GATACCTGCTTACCTGGAATACAGATTAAGATATTTTGGACACATTCGAGAGAGAA 1980  
Db 2143 GATACCTGCTTACCTGGAATACAGATTAAGATATTTTGGACACATTCGAGAGAGAA 2202  
QY 1981 CTTTGAAGCCGCTGTCATCTAGCATTTGAGGAATTTCTTCTCTTTGAAAGTTGTC 2040  
Db 2203 CTTTGAAGCCGCTGTCATCTAGCATTTGAGGAATTTCTTCTCTTTGAAAGTTGTC 2262  
QY 2041 ACTACTTAG 2049  
Db 2263 ACTACTTAG 2271

RESULT 3  
AAA97632 standard; cDNA: 3108 BP.  
AAA97632:  
AA97632:  
02-FEB-2001 (first entry)  
Mouse neutral/alkaline ceramidase clone, SEQ ID NO:11.  
Neutral/alkaline ceramidase: mouse; murine; recombinant production;  
cellular ceramide content control; antibody; membrane synthesis;  
lipid engineering; ceramide metabolism; drug development; ss.  
Mus sp.  
WO200058448-A1.  
05-OCT-2000.  
24-MAR-2000; 2000WO-JP01802.  
26-MAR-1999; 99JP-0084743.

XX (TAKI ), TAKARA SHUZO CO LTD.  
PA Ito M;  
XX WPI: 2000-619079/59.  
DR Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
XX reagent for studying structure and functions of ceramide as well as  
XX developing drugs for diseases associated with ceramide metabolism  
PS Example 4; Page 55-56; 76pp; Japanese.  
XX The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents a clone encoding  
XX mouse neutral/alkaline ceramidase.  
SQ Sequence 3108 BP; 888 A; 760 C; 714 G; 746 T; 0 other:  
Query Match 100.0%; Score 2049; DB 21; Length 3108;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTCAGTGGCTTACTACATTTGGCGTTGGAGAGCGGATTTCACAGCAAGTGTACATATTC 60  
Db 948 TTCAGTGGCTTACTACATTTGGCGTTGGAGAGCGGATTTCACAGCAAGTGTACATATTC 1007  
QY 61 AATTGATGGGCTATGGCAAAATGGCCAGATGACAGGGGTCTCTCCACAGGCTGTC 120  
Db 1008 AATTGATGGGCTATGGCAAAATGGCCAGATGACAGGGGTCTCTCCACAGGCTGTC 1067  
QY 121 AGCGTGCTTTATCTTGGCGATCCAGATGGGTCAATTCGAAATGGCATTTGTGAGCGTG 180  
Db 1068 AGCGTGCTTTATCTTGGCGATCCAGATGGGTCAATTCGAAATGGCATTTGTGAGCGTG 1127  
QY 181 GACATATATGATTTCCCAACGACTGAGTGGTGGAGTCTCGAAGACTAGAGATTA 240  
Db 1128 GACATATATGATTTCCCAACGACTGAGTGGTGGAGTCTCGAAGACTAGAGATTA 1187  
QY 241 TATGGCTCTGTATCGAAGAGACATGTTATCCGAGTGGCATTCACACACACTCTGGC 300  
Db 1188 TATGGCTCTGTATCGAAGAGACATGTTATCCGAGTGGCATTCACACACACTCTGGC 1247  
QY 301 CCAGCAGGGTTTTCCAATATACACTATATCTCGCCAGCGAGGATTCAGCAACCGG 360  
Db 1248 CCAGCAGGGTTTTCCAATATACACTATATCTCGCCAGCGAGGATTCAGCAACCGG 1307  
QY 361 ACCTTTCAGTACATAGTCTCTGGGATCATGAAGAGACTTGTATAGCTACACAAATCTT 420  
Db 1308 ACCTTTCAGTACATAGTCTCTGGGATCATGAAGAGACTTGTATAGCTACACAAATCTT 1367  
QY 421 AACGAGCAAAATCTTATCAACAAAGGAATGTTGCTAATGTGCAGATCAACGAGG 480  
Db 1368 AACGAGCAAAATCTTATCAACAAAGGAATGTTGCTAATGTGCAGATCAACGAGG 1427  
QY 481 CCCTCCTTACCTTGTGAATCCAGAGTCCAGAGAGAGCAAGATTTCTTCAACACAGAG 540  
Db 1428 CCCTCCTTACCTTGTGAATCCAGAGTCCAGAGAGAGCAAGATTTCTTCAACACAGAG 1487  
QY 541 AAGGAATGCTGTCTTGAACCTGTGATTTGAATGGAAGAGACTTGGCTTATTCAGC 600  
Db 1488 AAGGAATGCTGTCTTGAACCTGTGATTTGAATGGAAGAGACTTGGCTTATTCAGC 1547  
QY 601 TGGTTTGCATCCACCCGCTGAGTGAACATATGCAACCACTTGTGAATAGTACAT 660

Db 1548 TGGTTTGCATCCACCCGTGAGCATGAACAATAGCAACCACTTTGTGAATAGTGAACAT 1607  
 QY 661 ATGGGCTATGCGGCTTACCTTTTGTGACAAGAAAAGAACAAAGGCTATCGCTGGACAG 720  
 Db 1608 ATGGGCTATGCGGCTTACCTTTTGTGACAAGAAAAGAACAAAGGCTATCGCTGGACAG 1667  
 QY 721 GGACGGTTTGTAGACAGGCTTTGCTTCATCAAAATCTCGAGAGCTGTCAACCAATCTT 780  
 Db 1668 GGACGGTTTGTAGACAGGCTTTGCTTCATCAAAATCTCGAGAGCTGTCAACCAATCTT 1727  
 QY 781 GGCCCGCATTTGTGTCAACACAGGGAGTCTTGTGACACGACAGACGACCTGTCCCAAC 840  
 Db 1728 GGCCCGCATTTGTGTCAACACAGGGAGTCTTGTGACACGACAGACGACCTGTCCCAAC 1787  
 QY 841 GGTGGGCTTAGCATGTGATGGCGACGGAGCCGGGCAAGACATGTTTAGAGCAACAC 900  
 Db 1788 GGTGGGCTTAGCATGTGATGGCGACGGAGCCGGGCAAGACATGTTTAGAGCAACAC 1847  
 QY 901 ATTATAGACGGATCATCTATCAGAGGCCAAGAGCTGTATGCCCTGTGCCCCAGAG 960  
 Db 1848 ATTATAGACGGATCATCTATCAGAGGCCAAGAGCTGTATGCCCTGTGCCCCAGAG 1907  
 QY 961 GTGACCGGCCAGCTGCTTGCAGTCAACAGTGGGTGACATGACAGATGTGACGCTCAG 1020  
 Db 1908 GTGACCGGCCAGCTGCTTGCAGTCAACAGTGGGTGACATGATGTGACGCTCAG 1967  
 QY 1021 CTCATGCGCACACACACAGTGAAGCGTAAACCGCCCTGGGCTACAGTTTGGCGCA 1080  
 Db 1968 CTCATGCGCACACACAGTGAAGCGTAAACCGCCCTGGGCTACAGTTTGGCGCA 2027  
 QY 1081 GGCACAAATTGATGAGTTTCGGGCGCTCAATATTACACAGGAACTACGAGGCGATCCA 1140  
 Db 2028 GGCACAAATTGATGAGTTTCGGGCGCTCAATATTACACAGGAACTACGAGGCGATCCA 2087  
 QY 1141 TTCTGGGACACTCTTCGGGACCAAGCTCTTGGGAAAACATCTGAGAGATTTAGAGTCT 1200  
 Db 2088 TTCTGGGACACTCTTCGGGACCAAGCTCTTGGGAAAACATCTGAGAGATTTAGAGTCT 2147  
 QY 1201 CAGAAAACCAACCAATCTGCTCAACAGTGGAGAGCTACAGATACCAATCTCTGGCAA 1260  
 Db 2148 CAGAAAACCAACCAATCTGCTCAACAGTGGAGAGCTACAGATACCAATCTCTGGCAA 2207  
 QY 1261 CCAGATATTGTTGATGATTCAGATTGTTTACCGTTGGGCTTGGCCATGCTATCCCT 1320  
 Db 2208 CCAGATATTGTTGATGATTCAGATTGTTTACCGTTGGGCTTGGCCATGCTATCCCT 2267  
 QY 1321 GGGGAATTAAACAACATGTCGGGACGACATTTCTGAGGCAATTAATAAAGATTGCA 1380  
 Db 2268 GGGGAATTAAACAACATGTCGGGACGACATTTCTGAGGCAATTAATAAAGATTGCA 2327  
 QY 1381 CTTTATGGATGAAGATATGACCGTGTGATTCGACAGCTCTAAGCAATTTTATACAT 1440  
 Db 2328 CTTTATGGATGAAGATATGACCGTGTGATTCGACAGCTCTAAGCAATTTTATACAT 2387  
 QY 1441 TACATTACACATATGAAGAATACAGGCTCAGCGGTACGAGCAGCATCTACAATCTAT 1500  
 Db 2388 TACATTACACATATGAAGAATACAGGCTCAGCGGTACGAGCAGCATCTACAATCTAT 2447  
 QY 1501 GGACCAACACACCTGTCTGACATACATCAACTCTTCAGAGCTCTGCTAAGCAATTTGCT 1560  
 Db 2448 GGACCAACACACCTGTCTGACATACATCAACTCTTCAGAGCTCTGCTAAGCAATTTGCT 2507  
 QY 1561 AGGACACAGTATGCAACATGAGCAGTGGTCCCGAGCCCACTTCTGAAAATCTAATA 1620  
 Db 2508 AGGACACAGTATGCAACATGAGCAGTGGTCCCGAGCCCACTTCTGAAAATCTAATA 2567  
 QY 1621 GCTTCACTTATCTCAATATGCGGATAGACACCAATTTGGCAAACTTTTGGGATGTC 1680  
 Db 2568 GCTTCACTTATCTCAATATGCGGATAGACACCAATTTGGCAAACTTTTGGGATGTC 2627  
 QY 1681 TTGACAGCCAGCAAAACCTGCAATACAGAGTGGGAGAAATGTTATTTTGTAGGC 1740  
 Db 2628 TTGACAGCCAGCAAAACCTGCAATACAGAGTGGGAGAAATGTTATTTTGTAGGC 2687

QY 1741 GCTAACCAAGATTCAGAGAGAACACAGACCATCAAACTTCTCACTGTGGAGAA 1800  
 Db 2688 GCTAACCAAGATTCAGAGAGAACACAGACCATCAAACTTCTCACTGTGGAGAA 2747  
 QY 1801 TACGAGACTCTGTAGCTGCTGCGAGATTAATGTATTAACAGTCCCTCTGGAGACGAG 1860  
 Db 2748 TACGAGACTCTGTAGCTGCTGCGAGATTAATGTATTAACAGTCCCTCTGGAGACGAG 2807  
 QY 1861 TTTTATTTGGCACAAGGAATCTGGGTCTGAGCAATGCAACATATATCTGGCATATCCA 1920  
 Db 2808 TTTTATTTGGCACAAGGAATCTGGGTCTGAGCAATGCAACATATATCTGGCATATCCA 2867  
 QY 1921 GATACGCTTACCTGCAATCTACAAATTAATTTTGGACACAAATGGAAGCAGAA 1980  
 Db 2868 GATACGCTTACCTGCAATCTACAAATTAATTTTGGACACAAATGGAAGCAGAA 2927  
 QY 1981 CTTCTGAAACCCGCTGCTCACTAGCATTTTGAAGAAATTTCTTCTCTTTGAAGTGTCT 2040  
 Db 2928 CTTCTGAAACCCGCTGCTCACTAGCATTTTGAAGAAATTTCTTCTCTTTGAAGTGTCT 2987  
 QY 2041 ACTACTTAG 2049  
 Db 2988 ACTACTTAG 2996  
 RESULT 4  
 AAA97635  
 ID AAA97635 standard; DNA; 4835 BP.  
 XX  
 AC AAA97635;  
 XX  
 DT 02-FEB-2001 (first entry)  
 DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:16.  
 XX  
 DE Neutral/alkaline ceramidase; mouse; murine; recombinant production;  
 XX cellular ceramide content control; antibody; membrane synthesis;  
 KW lipid engineering; ceramide metabolism; drug development; ds.  
 XX  
 OS Mus sp.  
 OS Synthetic.  
 OS  
 PN MO200058448-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 24-MAR-2000; 2000MO-JP01802.  
 PF  
 XX 26-MAR-1999; 99JP-0084743.  
 PR  
 XX (TAKI ) TAKARA SHUZO CO LTD.  
 PA  
 XX Ito M;  
 PI  
 XX MPI; 2000-619079/59.  
 DR  
 XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
 PT reagent for studying structure and functions of ceramide as well as  
 PT developing drugs for diseases associated with ceramide metabolism  
 XX  
 PS Example 6; Page 67-70; 76pp; Japanese.  
 CC The invention relates to a mouse neutral/alkaline ceramidase and to  
 CC nucleic acids encoding it. The invention also relates to expression  
 CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
 CC ceramidase, the recombinant production of the ceramidase, an antibody  
 CC against the ceramidase, and a method for controlling the ceramide  
 CC content in cells and/or tissues by introducing the gene or its antisense  
 CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
 CC as a lipid engineering reagent for studying structure and functions of  
 CC ceramide as well as developing drugs for diseases associated with  
 CC ceramide metabolism. The present sequence represents a clone encoding

CC mouse neutral/alkaline ceramidase.

**SQ** Sequence 4835 BP; 1460 A; 996 C; 1048 G; 1331 T; 0 other;

Query Match	98.48;	Score 2016;	DB 21;	Length 4835;
-------------	--------	-------------	--------	--------------

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2016; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	34	AATTGGACAGACAAAGTGTCAGATATCAATTGATGGGCTATGGCAAAAATAGGCGAGAT	93
Db	696	GATTTCACAGACAAAGTGTCAGATATCAATTGATGGGCTATGGCAAAAATAGGCGAGAT	755
OY	94	GCACGGGGCTCTCTACCAGGCTGTTCAGCCGTCCTTTATCTTGGCGGATCCAGATGGG	153
Db	756	GCACGGGGCTCTCTACCAGGCTGTTCAGCCGTCCTTTATCTTGGCGGATCCAGATGGG	815
OY	154	TCAAAATCGAATGGCAATTTGTGAGGGTGGAACTATGTATGATTTCCCAACGACTGAGTTG	213
Db	816	TCAAAATCGAATGGCAATTTGTGAGGGTGGAACTATGTATGATTTCCCAACGACTGAGTTG	875
OY	214	GAGGTCCCGAAGAGACTAGAGATGAATATGGCTCTCTGTTCGAAGAGACAATGTTATC	273
Db	876	GAGGTCCCGAAGAGACTAGAGATGAATATGGCTCTCTGTTCGAAGAGACAATGTTATC	935
OY	274	CTGAGTGGCAATTCACACACTCTGGCCACAGAGGGTTTTCATATATACACTGTATATA	333
Db	936	CTGAGTGGCAATTCACACACTCTGGCCACAGAGGGTTTTCATATATACACTGTATATA	995
OY	334	CTGCCCGACGAGGGATTCAGCAACCGGACCTTTCAGTACATAGTCTCGGGATCATGAAG	393
Db	996	CTGCCCGACGAGGGATTCAGCAACCGGACCTTTCAGTACATAGTCTCGGGATCATGAAG	1055
OY	394	AGCAATTGATATAGCTCACACAAATCTTTAAACAGGCAAAATCTTTATCAACAAAGGAAT	453
Db	1056	AGCAATTGATATAGCTCACACAAATCTTTAAACAGGCAAAATCTTTATCAACAAAGGAAT	1115
OY	454	GTTCTCTAAATGTGCAGATCAACCGGAAGCCCTCTCTTACCTTGTGAATCCACACTGACAG	513
Db	1116	GTTCTCTAAATGTGCAGATCAACCGGAAGCCCTCTCTTACCTTGTGAATCCACACTGACAG	1175
OY	514	AGACCAAGGTATCTTCTTCAACACAGACAGGAAGAAATGCGTGTGAAACTGTGGATTTTG	573
Db	1176	AGACCAAGGTATCTTCTTCAACACAGACAGGAAGAAATGCGTGTGAAACTGTGGATTTTG	1235
OY	574	AATGGAGAAAGCTTTGGGCTTATAGAGCTGTTCGCATCCACCCCGTAGCATGAACAT	633
Db	1236	AATGGAGAAAGCTTTGGGCTTATAGAGCTGTTCGCATCCACCCCGTAGCATGAACAT	1295
OY	634	AGCAACCACTTTTGTGAATATGTGACAAATATGGGCTATGCGGCTTACCTTTTGTAGCAAGAA	693
Db	1296	AGCAACCACTTTTGTGAATATGTGACAAATATGGGCTATGCGGCTTACCTTTTGTAGCAAGAA	1355
OY	694	AAGAAACAAAGGCTATGCGCCTGAGACAGGAGACGTTTGTAGAGAGCTTTGCTTCATCAAT	753
Db	1356	AAGAAACAAAGGCTATGCGCCTGAGACAGGAGACGTTTGTAGAGAGCTTTGCTTCATCAAT	1415
OY	754	CTCGGAGACGTGTACCCCAACATCTTGGCCGCAATTTGTCAACACAGGGAGACTTGT	813
Db	1416	CTCGGAGACGTGTACCCCAACATCTTGGCCGCAATTTGTCAACACAGGGAGACTTGT	1475
OY	814	GACAAACGACAAGACACCTGTGCCAACCGTGGGCTTAGCATGTGCATGGCCAGCGACCT	873
Db	1476	GACAAACGACAAGACACCTGTGCCAACCGTGGGCTTAGCATGTGCATGGCCAGCGACCT	1535
OY	874	GGACAAGACATGTTTGGAGAGCACACACTTATAGGAGGGATCATCTATCAAAAGGCCAAG	933
Db	1536	GGACAAGACATGTTTGGAGAGCACACACTTATAGGAGGGATCATCTATCAAAAGGCCAAG	1595
OY	934	GAGCTGTATGCTCTCGCCTCCACAGAGATGACCGGCCAGTGTTGCAAGCTCACACAGTGG	993
Db	1596	GAGCTGTATGCTCTCGCCTCCACAGAGATGACCGGCCAGTGTTGCAAGCTCACACAGTGG	1655
OY	994	GTGAACATGACAGATGTGAGGCTCAGCTCAATGCAACACAGAGTGAAGACGTGTATA	1053

Db	1656	GTGAAACATGACAGATGTGAGCGTCCAGCTCATGTCACACACACAGTGAAGACGTGTAAA	1715
OY	1054	CTGCGCCCTGGGCTACAGTTTGTGCGGACGACCAATTTGATGAGAGTTTGGGCGCTCAATATT	1113
Db	1716	CTGTCCCTGGGCTACAGTTTGTGCGGACGACCAATTTGATGAGAGTTTGGGCGCTCAATATT	1775
OY	1114	ACACAGGGAACCTACGGAAAGGGGATCCAAPTGCGGACACTGTGGGACCAAGCTTTGGGA	1173
Db	1776	ACACAGGGAACCTACGGAAAGGGGATCCAAPTGCGGACACTGTGGGACCAAGCTTTGGGA	1835
OY	1174	AAACCATCTGTAAGAGATTTGTAGAGTGTGAGAAACCCAAACCAATCTGCTTACAGTGA	1233
Db	1836	AAACCATCTGTAAGAGATTTGTAGAGTGTGAGAAACCCAAACCAATCTGCTTACAGTGA	1895
OY	1234	GAGCTGACGATACCAACATCCCTGGCAACGATTTTGTATGTTGACATTTGTATCCGTT	1293
Db	1896	GAGCTGACGATACCAACATCCCTGGCAACGATTTTGTATGTTGACATTTGTATCCGTT	1955
OY	1294	GGGTCTTGTGGCATAGCTGCTATTCCCTGGGGAATTAACAACATGTGCGGACGAAGATTT	1353
Db	1956	GGGTCTTGTGGCATAGCTGCTATTCCGCGGGAAATTAACAACCATGTGCGGACGAAGATTT	2015
OY	1354	CGTAGGCAATTAATAAAGAATTTGGCACTTTATGGATGAAGATATAGCCGTGTATTTC	1413
Db	2016	CGTAGGCAATTAATAAAGAATTTGGCACTTTATGGATGAAGATATAGCCGTGTATTTC	2075
OY	1414	GCAGGTCTAAGCAATGTTTATATACACATTACATTAATACCATATGAAGATATACAGGCTCAG	1473
Db	2076	GCAGGTCTAAGCAATGTTTATATACACATTACATTAATACCATATGAAGATATACAGGCTCAG	2135
OY	1474	CGGTACGAGCGAGCATCTACAACTTATGAGACACACACCCGTGTGCATACATCCAACTC	1533
Db	2136	CGGTACGAGCGAGCATCTACAACTTATGAGACACACACCCGTGTGCATACATCCAACTC	2195
OY	1534	TTTCAGAGACCTCTGTAAGGCAATTTGCTACGAGACACATAGCCAATGAGACAGTGTGCC	1593
Db	2196	TTTCAGAGACCTCTGTAAGGCAATTTGCTACGAGACACACATAGCCAATGAGACAGTGTGCC	2255
OY	1594	GAGCCTCCATTTCTCAAAAATCTAATAGCTTCACTTATTCCTAATATTGGGATAGAGCA	1653
Db	2256	GAGCCTCCATTTCTCAAAAATCTAATAGCTTCACTTATTCCTAATATTGGGATAGAGCA	2315
OY	1654	CCAATTTGGCAAAACTTTTGGGGATGTCTTGAGCGACAGCAAAAACCTGAAATACAGAGTGGGA	1713
Db	2316	CCAATTTGGCAAAACTTTTGGGGATGTCTTGAGCGACAGCAAAAACCTGAAATACAGAGTGGGA	2375
OY	1714	GAACTGTTGAAGTTATATTATTGTAGCGGCTAACCCAAAGAAATTCAGCAGAGAACACACC	1773
Db	2376	GAACTGTTGAAGTTATATTATTGTAGCGGCTAACCCAAAGAAATTCAGCAGAGAACACACC	2435
OY	1774	CATCAAAACCTTCCTACCTGTGGAAAAATACGAGAGACTGTGAGCTGAGCTGGCAGATPATG	1833
Db	2436	CATCAAAACCTTCCTACCTGTGGAAAAATACGAGAGACTGTGAGCTGAGCTGGCAGATPATG	2495
OY	1834	TATTAACGATGCTCTCTGGGAGACGAGGTTTATTGGGACAAAGGAATACTGGGCTGAGC	1893
Db	2496	TATTAACGATGCTCTCTGGGAGACGAGGTTTATTGGGACAAAGGAATACTGGGCTGAGC	2555
OY	1894	AATGCAACAAATATATCTGGCATATTTCCAGATATCTGCCATACCTGGAAATCTACAGAAATAGA	1953
Db	2556	AATGCAACAAATATATCTGGCATATTTCCAGATATCTGCCATACCTGGAAATCTACAGAAATAGA	2615
OY	1954	TATTTTGGACACAATCGGAAGCAGAGAACTTCTGAACCCCGCTGTCAATAGACTATTGAA	2013
Db	2616	TATTTTGGACACAATCGGAAGCAGAGAACTTCTGAACCCCGCTGTCAATAGACTATTGAA	2675
OY	2014	GGAATTTCTCTCTCTTTTGAAGTTGCACTACTTAG 2049	
Db	2676	GGAATTTCTCTCTCTTTTGAAGTTGCACTACTTAG 2711	

## RESULT 5

AAH98033  
ID AAH98033 standard; DNA: 551 BP.  
XX  
AC AAH98033:  
XX  
DT 10-OCT-2001 (first entry)  
XX  
DE Murine 7-transmembrane G-protein coupled receptor coding sequence #277.  
XX  
KM Murine: stromal stem cell; signalling; vaccine; 7TM-GPCR;  
XX 7-transmembrane G-protein coupled protein receptor; ds.  
XX  
OS Mus sp.  
XX MO200160999-A1.  
XX  
XX 23-AUG-2001.  
XX  
PF 14-FEB-2001; 2001MO-0504700.  
XX  
PR 14-FEB-2000; 2000US-0182377.  
XX  
PA (IMCL-) IMCLONE SYSTEMS INC.  
XX (UYPR-) UNIV PRINCETON.  
XX  
PI Lemischka IR, Witte L, Pereira DS;  
XX  
XX WPI; 2001-522596/57.  
XX  
DR DNA Sequences encoding 7-transmembrane G-protein coupled protein  
PT receptors characteristic of hematopoietic stem cells, useful for  
PT treating leukemia -  
XX  
PS Claim 1: Page 125; 176pp; English.  
XX  
XX The present invention relates to murine coding sequences for  
CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The  
CC present sequence is one such murine 7TM-GPCR coding sequence. The present  
CC sequence was derived from stromal stem cells. The present sequence  
CC and its corresponding protein are useful in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate 7TM-GPCR expression.  
CC 7TM-GPCRs identify specific signalling molecules, to activate an  
CC effector-signalling cascade that triggers an intracellular response and  
CC eventually a biological effect.  
XX  
SQ Sequence 551 BP; 145 A; 135 C; 150 G; 120 T; 1 other:  
Query Match 16.1%; Score 329; DB 22; Length 551;  
Best Local Similarity 99.8%; Pred. No. 6.1e-156;  
Matches 449; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 802 GGGGACTCTTGTGACACAGACAGACAGACAGACAGAGGCTGAGTATGATGATG 861  
DB 1 GGGGACTCTTGTGACACAGACAGACAGACAGACAGAGGCTGAGTATGATGATG 60  
QY 862 GCCAGCGACCTGACAGACAGATGTTGAGAGACACACATTTAGAGAGATCATCTAT 921  
DB 61 GCCAGCGACCTGACAGACAGATGTTGAGAGACACACATTTAGAGAGATCATCTAT 120  
QY 922 CAGAGGCCAAGAGAGCTGTATGCTCTGCTCCAGAGAGTGACCGCCAGTGTGCA 981  
DB 121 CAGAGGCCAAGAGAGCTGTATGCTCTGCTCCAGAGAGTGACCGCCAGTGTGCA 180  
QY 982 GCTCACCAGTGGGTGAAATGACAGATGAGAGGCTCCAGTCAATGCCACACACAGTG 1041  
DB 181 GCTCACCAGTGGGTGAAATGACAGATGAGAGGCTCCAGTCAATGCCACACACAGTG 240  
QY 1042 AAGACGTGTAAACCTGCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1101  
DB 241 AAGACGTGTAAACCTGCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 299  
QY 1102 GGCCTCAATTTACACAGGAGACTACGGAAGGAGATTCATTTGGAGACCTCTCGGAC 1161

DB 300 GGCCTCAATTTACACAGGAGACTACGGAAGGAGATTCATTTGGAGACCTCTCGGAC 359  
QY 1162 CAGCTCTTGGGAAACCATCTGAGAGATTTGAGAGTGTGAGAGAACCCAAATCCTG 1221  
DB 360 CAGCTCTTGGGAAACCATCTGAGAGATTTGAGAGTGTGAGAGAACCCAAATCCTG 419  
QY 1222 CTTACAGTGTGAGAGCTGACATACACAT 1251  
DB 420 CTTACAGTGTGAGAGCTGACATACACAT 449  
RESULT 6  
AAF9977  
ID AAF9977 standard; DNA: 2283 BP.  
XX  
XX AAF9977;  
XX  
XX 20-JUL-2001 (first entry)  
XX  
DE DNA encoding rat protein involved in acid amide bond hydrolysis.  
XX  
XX Rat; acid amide bond hydrolysis; ceramide; cytosolic; cancer; ds.  
XX  
OS Rattus norvegicus.  
XX  
XX Key location/Qualifiers  
FH CDS 1..2283  
FT /\*tag- a  
FT /partial  
PN /note="this sequence does not contain a stop codon"  
XX  
XX JP2001057890-A.  
XX  
XX 06-MAR-2001.  
XX  
XX 23-AUG-1999; 99JP-0235218.  
XX  
XX 23-AUG-1999; 99JP-0235218.  
XX  
PA (MITU) MITSUBISHI CHEM CORP.  
XX  
XX WPI; 2001-304133/32.  
DR P-PSDB: AAB97029.  
DR  
XX Novel protein, used to treat abnormal cell growth e.g. in cancers, is  
PT capable of hydrolyzing the acid amide bond between sphingosine and the  
PT fatty acid in ceramide -  
XX  
PS Claim 3; Page 9-11; 14pp; Japanese.  
XX  
XX The present sequence is provided in a specification relating to a  
CC protein having a 761 residue amino acid sequence or its variant  
CC comprising a replacement, deletion, insertion, addition or reversion of  
CC at least one amino acid, but retaining the activity of hydrolyzing the  
CC acid amide bond between sphingosine and the fatty acid in a ceramide.  
CC The protein can be used to treat diseases caused by abnormal cell  
XX growth such as cancers.  
SQ Sequence 2283 BP; 641 A; 584 C; 524 G; 534 T; 0 other:  
Query Match 3.3%; Score 68; DB 22; Length 2283;  
Best Local Similarity 100.0%; Pred. No. 8.3e-24;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1057 GCCCTGGGCTACAGTTTGGCCGACGACCAATTTGAGTGTGAGGCTCAATATTACA 1116  
DB 1294 GCCCTGGGCTACAGTTTGGCCGACGACCAATTTGAGTGTGAGGCTCAATATTACA 1353  
QY 1117 CAGGGAAC 1124  
DB 1354 CAGGGAAC 1361



RESULT 7  
ABN53673  
ID ABN53673 standard; DNA; 65 BP.  
XX  
XX ABN53673;  
AC  
XX  
XX 15-JUL-2002 (first entry)  
DT  
XX  
XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:26421.  
DE  
XX  
XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
KW splice variant; transcriptome; oligonucleotide library; ss.  
XX  
XX Mus musculus.  
OS  
XX WO200210449-A2.  
PN  
XX  
XX 07-FEB-2002.  
PD  
XX  
XX 20-JUL-2001; 2001WO-1B01903.  
PF  
XX  
XX 28-JUL-2000; 2000US-221607P.  
PR  
XX 02-MAY-2001; 2001US-287724P.  
XX  
XX (COMP-) COMPUGEN INC.  
PA  
XX  
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
PI  
XX  
XX WPI; 2002-257383/30.  
DR  
XX  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes  
XX  
XX  
XX Example 1; SEQ ID 26421; 47pp; English.  
PS  
XX  
XX The present invention describes oligonucleotide libraries for detecting  
CC messenger RNAs that populate a (sub-)transcriptome, where the  
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
CC transcription units that populate a genome. The library comprises  
CC several oligonucleotides, each capable of hybridizing selectively to a  
CC set of messenger RNAs transcribed from a given transcription unit of  
CC the genome, which encodes one or more messenger RNA splice variants.  
CC The oligonucleotide libraries are useful for detecting mRNAs from a  
CC biological sample, in expression profiling studies, in qualitatively or  
CC quantitatively characterizing the corresponding transcriptome, and in  
CC detecting RNA transcripts and splice variants of human or animal  
CC transcriptomes. The libraries may also be used as specialised mini  
CC libraries to detect transcripts of a sub-transcriptome under a  
CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN27253 to ABN59389 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_Pct\_sequences.  
XX  
XX  
SQ Sequence 65 BP; 20 A; 18 C; 14 G; 13 T; 0 other;  
Query Match 3.2%; Score 65; DB 24; Length 65;  
Best Local Similarity 100.0%; Pred. No. 2.8e-22;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1767 CCAGACCCATCAAAACCTTCCTCAGTGTGAGAAATACGAGACTGTAGCTGACTGGCA 1826  
DB 1 CCAGACCCATCAAAACCTTCCTCAGTGTGAGAAATACGAGACTGTAGCTGACTGGCA 60  
OY 1827 GATTA 1831

DB 61 GATTA 65  
RESULT 8  
AAF99979  
ID AAF99979 standard; CDNA; 539 BP.  
XX  
XX AAF99979;  
AC  
XX  
XX 20-JUL-2001 (first entry)  
DT  
XX  
XX PCR product isolated from rat CDNA.  
DE  
XX  
XX Rat; acid amide bond hydrolysis; ceramide; cytosolic; cancer; ss.  
XX  
XX Rattus norvegicus.  
OS  
XX JP2001057890-A.  
PN  
XX  
XX 06-MAR-2001.  
PD  
XX  
XX 23-AUG-1999; 99JP-0235218.  
PF  
XX  
XX 23-AUG-1999; 99JP-0235218.  
PR  
XX  
XX (MITU ) MITSUBISHI CHEM CORP.  
PA  
XX  
XX WPI; 2001-304133/32.  
DR  
XX  
XX  
XX Novel protein, used to treat abnormal cell growth e.g. in cancers, is  
PT capable of hydrolyzing the acid amide bond between sphingosine and the  
PT fatty acid in ceramide  
XX  
XX  
XX Example 3; Page 14; 14pp; Japanese.  
PS  
XX  
XX The present sequence is provided in a specification relating to a  
CC protein having a 761 residue amino acid sequence or its variant  
CC comprising a replacement, deletion, insertion, addition or reversion of  
CC at least one amino acid, but retaining the activity of hydrolysing the  
CC acid amide bond between sphingosine and the fatty acid in a ceramide.  
CC The protein can be used to treat diseases caused by abnormal cell  
CC growth such as cancers.  
XX  
XX  
SQ Sequence 539 BP; 181 A; 110 C; 113 G; 135 T; 0 other;  
Query Match 2.6%; Score 53; DB 22; Length 539;  
Best Local Similarity 100.0%; Pred. No. 3.3e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1765 AACGAGACCCATCAAAACCTTCCTCAGTGTGAGAAATACGAGACTGTAGC 1817  
DB 154 AACGAGACCCATCAAAACCTTCCTCAGTGTGAGAAATACGAGACTGTAGC 206  
RESULT 9  
AAD22668  
ID AAD22668 standard; CDNA; 2507 BP.  
XX  
XX AAD22668;  
AC  
XX  
XX 26-FEB-2002 (first entry)  
DT  
XX  
XX Human ceramidase CDNA.  
DE  
XX  
XX Human; ceramidase; nephrotropic; antipsoriatic; antiscars therapy;  
KW gene therapy; proliferative disorder; cancer; cardiovascular disease;  
KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;  
KW neurotropic; signal transduction; breast cancer; autoimmune disorder;  
KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;  
KW glomerular disease; ss.  
XX  
XX Homo sapiens.



XX Key Location/Qualifiers  
FH CDS 39..2324  
FT /tag= a  
FT /product= "Human ceramidase protein"  
FT /note= "There is an additional stop codon from position 2325-2327"  
FT 39..95  
FT s1g-peptide  
FT mat-peptide 96..2321  
FT /tag= b  
FT /tag= c  
FT /product= "Mature ceramidase protein"  
XX WO200155410-A2.  
XX 02-AUG-2001.  
XX PD  
XX 29-JAN-2001; 2001WO-US02866.  
XX PE  
XX 28-JAN-2000; 2000US-178975P.  
XX PR  
XX (MUSC-) MUSC FOUND RES DEV.  
XX PA  
XX Hannun YA, El Bawab S;  
XX PI  
XX WPI: 2002-025687/03.  
XX DR P-PSDB: AAE13544.  
XX PT Human mitochondrial ceramidase protein and gene, modulation of which is useful for preventing and treating proliferative disorders e.g. cancer, cardiovascular disease, inflammation and neurodegenerative disorders -  
XX PS Claim 1; Fig 3; 113pp; English.  
XX The present sequence is a cDNA encoding human mitochondrial ceramidase protein. Ceramidase protein and gene are useful for treating a disease or disorder involving cell over proliferation or sphingolipid signal transduction especially breast cancer, cardiovascular disorder or inflammation. Ceramidase protein and gene are useful for treating disorders involving deficient cell proliferation or growth e.g. neurodegenerative disorders (Alzheimer's disease), growth deficiencies and lesions. Ceramidase protein is also useful for diagnosis of hyperproliferative diseases. Ceramidase gene can be used as an immunogen to generate antibodies which are useful for diagnosis, prevention and treatment of hyperproliferative diseases and for detecting ceramidase gene product in a biological sample. The hyperproliferative disorders include cancers and autoimmune disorders such as lupus nephritis, glomerular disease. Ceramidase gene is also useful in antisense therapy.  
XX Sequence 2507 BP; 720 A; 583 C; 567 G; 637 T; 0 other:  
SQ  
Query Match 2.3%; Score 47; DB 24; Length 2507;  
Best Local Similarity 100.0%; Pred. No. 3.5e-13;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

KW human immunodeficiency virus; HIV; infection; immunological disease;  
KW gene therapy; gene; ss.  
XX Homo sapiens.  
XX OS  
XX Key Location/Qualifiers  
FH CDS 1312..1641  
FT /tag= a  
FT /product= "sialyl transferase (HST30-1) 11.99"  
FT /note= "Claimed in claim 6"  
XX WO200233075-A1.  
XX 25-APR-2002.  
XX PD  
XX 25-JUN-2001; 2001WO-CN01058.  
XX PF  
XX 28-JUN-2000; 2000CN-0116809.  
XX PR  
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
XX PA  
XX Mao Y, Xie Y;  
XX PI  
XX WPI: 2002-340235/37.  
XX DR P-PSDB: ABB77549.  
XX PT Human sialyl transferase (HST30-1) 11.99 and encoding polynucleotide, used in diagnosis and treatment of malignant tumors, hemopathy, human immunodeficiency virus infection, immunological diseases and inflammation -  
XX PS Claim 6; Page 31-33; 39pp; Chinese.  
XX The invention relates to human sialyl transferase (HST30-1) 11.99 with cytostatic, virucidal, immunomodulatory, antiinflammatory and haemostatic activity. The protein and encoding polynucleotide are used in diagnosis and treatment of malignant tumor, hemopathy, human immunodeficiency virus (HIV) infection, immunological diseases and various inflammations. The polynucleotide is useful in gene therapy.  
XX Sequence 3058 BP; 932 A; 544 C; 597 G; 985 T; 0 other:  
SQ  
Query Match 2.1%; Score 43; DB 24; Length 3058;  
Best Local Similarity 100.0%; Pred. No. 3.8e-11;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1937 GAATCTACAGATAAGATATTTTGGACACAAATCGAGCAGCA 1979  
DB 519 GAATCTACAGATAAGATATTTTGGACACAAATCGAGCAGCA 561

RESULT 11  
AA97627  
ID AAA97627 standard; DNA: 38 BP.  
XX AC  
XX AAA97627:  
XX 02-FEB-2001 (first entry)  
XX DT  
XX DE Mouse neutral/alkaline ceramidase PCR primer, SEQ ID NO:6.  
XX KW Neutral/alkaline ceramidase; mouse; murine; recombinant production; cellular ceramide content control; antibody; membrane synthesis;  
XX KM Lipid engineering; ceramide metabolism; drug development; PCR primer; ss.  
XX OS Mus sp.  
XX WO200058448-A1.  
XX PN  
XX 05-OCT-2000.  
XX PD  
XX 24-MAR-2000; 2000WO-JP01802.  
XX PF  
XX Human; sialyl transferase; HST30-1; enzyme; cytostatic; virucidal;  
KW immunomodulatory; antiinflammatory; haemostatic; malignant tumour;

PR 26-MAR-1999; 99JP-0084743.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX Ito M;  
XX  
XX WPI; 2000-619079/59.  
XX  
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
PT reagent for studying structure and functions of ceramide as well as  
PT developing drugs for diseases associated with ceramide metabolism  
XX  
XX Example 3; Page 53; 76pp; Japanese.  
XX  
XX The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents a PCR primer  
CC used to isolate nucleic acids encoding mouse neutral/alkaline  
CC ceramidase.  
XX  
XX Sequence 38 BP; 8 A; 11 C; 9 G; 10 T; 0 other;  
SQ  
Query Match 1.9%; Score 38; DB 21; Length 38;  
Best Local Similarity 100.0%; Pred. No. 1.3e-08;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 733 GCAGGCTTGCTTCATCAATCTCGAGACGTGCACC 770  
DB 1 GCAGGCTTGCTTCATCAATCTCGAGACGTGCACC 38  
RESULT 12  
AAC03574  
ID AAC03574 standard; CDNA; 299 BP.  
XX  
XX AAC03574;  
AC  
XX 06-OCT-2000 (first entry)  
DT  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 3572.  
DE  
XX  
XX Human: 5' EST: expressed sequence tag; secreted protein; CDNA isolation;  
KM gene therapy; chromosome mapping; ss.  
KM  
XX Homo sapiens.  
OS  
XX EP1033401-A2.  
PN  
XX 06-SEP-2000.  
PD  
XX 21-FEB-2000; 2000EP-0200610.  
PE  
XX 26-FEB-1999; 99US-0122487.  
PR  
XX  
XX (GEST ) GENSET.  
PA  
XX Dumas Malne Edwards J, Duclet A, Giordano J;  
PI  
XX WPI; 2000-500381/45.  
DR  
XX P-PSDB; AAC03568.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 3572; 71pp + CD-ROM; English.

XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC cDNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
XX Sequence 299 BP; 84 A; 68 C; 76 G; 68 T; 3 other;  
SQ  
Query Match 1.5%; Score 31; DB 21; Length 299;  
Best Local Similarity 100.0%; Pred. No. 4.5e-05;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1682 TGCAGCCAGCAAAACCTGATACAGAGTGGC 1712  
DB 264 TGCAGCCAGCAAAACCTGATACAGAGTGGC 294  
RESULT 13  
AAZ06239  
ID AAZ06239 standard; DNA; 1194 BP.  
XX  
XX AAZ06239;  
AC  
XX 30-SEP-1999 (first entry)  
DT  
XX  
XX Human secreted protein gene No. 21.  
DE  
XX  
XX Human: secreted protein; fusion protein; gene therapy; protein therapy;  
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;  
KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KM immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KM inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;  
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
XX Homo sapiens.  
OS  
XX WO9935158-A1.  
PN  
XX 15-JUL-1999.  
PD  
XX 06-JAN-1999; 99WO-US00108.  
PE  
XX 07-JAN-1998; 98US-0070704.  
PR 07-JAN-1998; 98US-0070657.  
PR 07-JAN-1998; 98US-0070658.  
PR 07-JAN-1998; 98US-0070692.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;  
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;  
XX  
XX WPI; 1999-444190/37.  
DR P-PSDB; AAY38406, AAY38464, AAY38465, AAY38466, AAY38467, AAY38468,  
DR AAY38469, AAY38470, AAY38471.  
XX  
XX New isolated human genes and the secreted polypeptides they encode  
PT  
XX Claim 1; Page 160; 227pp; English.  
XX  
XX This sequence represents a nucleic acid molecule which encodes a

CC secreted human protein. The gene number is given in the descriptor line.  
CC The gene can be used to generate fusion proteins by linking to the gene  
CC to a human immunoglobulin Fc portion (e.g. AA06210) for increasing the  
CC stability of the fused protein as compared to the human protein only.  
CC The invention relates to 36 novel genes and their fragments (nucleic  
CC acid sequences: AA206219-206263; amino acid sequences AAY8386-Y38498)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 36  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AA206219 for described uses).

XX SQ Sequence 1194 BP; 358 A; 291 C; 249 G; 296 T; 0 other;

Query Match 1.4%; Score 29; DB 20; Length 1194;  
Best Local Similarity 100.0%; Pred. No. 0.00046;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 CAGATATCAATTGATGGCTATGGCAA 81  
DB 478 CAGATATCAATTGATGGCTATGGCAA 506  
|||||

RESULT 14

AA22672 ID AAD22672 standard; DNA; 38 BP.

AC AAD22672;

DT 26-FEB-2002 (first entry)

DE Human ceramidase cDNA cloning reverse PCR primer #2.

XX Human; ceramidase; nephrotropic; antipsoriatic; antitense therapy;  
XX gene therapy; proliferative disorder; cancer; cardiovascular disease;  
XX inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;  
XX neurotropic; signal transduction; breast cancer; autoimmune disorder;  
XX Alzheimer's disease; growth deficiency; lesion; lupus nephritis;  
XX glomerular disease; PCR primer; ss.

OS Homo sapiens.

PN WO200155410-A2.

XX 02-AUG-2001.

PF 29-JAN-2001; 2001WO-US02866.

XX 28-JAN-2000; 2000US-178975P.

XX (MUSC-) MUSC FOUND RES DEV.

PI Hannun YA, El Bawab S;

DR WPI; 2002-025687/03.

XX Human mitochondrial ceramidase protein and gene, modulation of which is  
XX useful for preventing and treating proliferative disorders e.g. cancer,  
XX cardiovascular disease, inflammation and neurodegenerative disorders -  
XX Example; Page 82; 113pp; English.

CC The present sequence is a PCR primer used for cloning human mitochondrial  
CC ceramidase DNA. Ceramidase protein and gene are useful for treating a  
CC disease or disorder involving cell over proliferation or sphingolipid  
CC signal transduction especially breast cancer, cardiovascular disorder or  
CC inflammation. Ceramidase protein and gene are useful for treating  
CC disorders involving deficient cell proliferation or growth e.g.  
CC neurodegenerative disorders (Alzheimer's disease), growth deficiencies  
CC and lesions. Ceramidase protein is also useful for diagnosis of  
CC hyperproliferative diseases. Ceramidase gene can be used as an immunogen

CC to generate antibodies which are useful for diagnosis, prevention and  
CC treatment of hyperproliferative diseases and for detecting ceramidase  
CC gene product in a biological sample. The hyperproliferative disorders  
CC include cancers and autoimmune disorders such as lupus nephritis,  
CC glomerular disease. Ceramidase gene is also useful in antitense therapy  
CC and gene therapy.

XX SQ Sequence 38 BP; 11 A; 10 C; 10 G; 7 T; 0 other;

Query Match 1.2%; Score 25; DB 24; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.05;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1682 TGCAGCCAGCAAACTGATACAG 1706  
DB 14 TGCAGCCAGCAAACTGATACAG 38  
|||||

RESULT 15  
ID AAA97636 standard; DNA; 24 BP.  
AC AAA97636;

DT 02-FEB-2001 (first entry)

DE Mouse neutral/alkaline ceramidase PCR primer, SEQ ID NO:17.

XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;  
XX cellular ceramidase content control; antibody; membrane synthesis;  
XX lipid engineering; ceramidase metabolism; drug development; PCR primer; ss.

OS Mus sp.

PN WO200058448-A1.

XX 05-OCT-2000.

PF 24-MAR-2000; 2000WO-JP01802.

XX 26-MAR-1999; 99JP-0084743.

XX (TAKI ) TAKARA SHUZO CO LTD.

PI Ito M;

DR WPI; 2000-619079/59.

XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
XX reagent for studying structure and functions of ceramidase as well as  
XX developing drugs for diseases associated with ceramidase metabolism -  
XX Example 7; Page 70; 76pp; Japanese.

CC The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramidase  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramidase as well as developing drugs for diseases associated with  
CC ceramidase metabolism. The present sequence represents a PCR primer  
CC used to amplify nucleic acids encoding mouse neutral/alkaline  
CC ceramidase.

XX SQ Sequence 24 BP; 8 A; 4 C; 6 G; 6 T; 0 other;

Query Match 1.2%; Score 24; DB 21; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY      885 GTTGGAGCAGACACATTATAGG 908
DB      1 GTTGGAGCAGCAGACATTATAGG 24

RESULT 16
AAA97637/C
ID      AAA97637 standard; DNA: 24 BP.
XX
XX
AC      AAA97637;
XX
XX
DT      02-FEB-2001 (first entry)
XX
XX
DE      Mouse neutral/alkaline ceramidase PCR primer, SEQ ID NO:18.
XX
XX
KM      Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KM      cellular ceramide content control; antibody; membrane synthesis;
XX      lipid engineering; ceramide metabolism; drug development; PCR primer; ss.
XX
OS      Mus sp.
XX
XX      WO200058448-A1.
XX
XX      05-OCT-2000.
XX
XX      24-MAR-2000; 2000MO-JP01802.
XX
XX      26-MAR-1999; 99JP-0084743.
XX
XX      (TAKI ) TAKARA SHUZO CO LTD.
XX
XX      Ito M;
PI
XX      WPI; 2000-619079/59.
XX
XX      Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT      reagent for studying structure and functions of ceramide as well as
PT      developing drugs for diseases associated with ceramide metabolism.
XX
XX      Example 7; Page 70; 76pp; Japanese.
XX
XX      The invention relates to a mouse neutral/alkaline ceramidase and to
CC      nucleic acids encoding it. The invention also relates to expression
CC      vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC      ceramidase, the recombinant production of the ceramidase, an antibody
CC      against the ceramidase, and a method for controlling the ceramide
CC      content in cells and/or tissues by introducing the gene or its antisense
CC      nucleic acid into the cells and/or tissues. The ceramidase can be used
CC      as a lipid engineering reagent for studying structure and functions of
CC      ceramide as well as developing drugs for diseases associated with
CC      ceramide metabolism. The present sequence represents a PCR primer
CC      used to amplify nucleic acids encoding mouse neutral/alkaline
CC      ceramidase.
XX
XX      Sequence 24 BP; 8 A; 7 C; 4 G; 5 T; 0 other;
SQ
OY      Query Match 1.28; Score 24; DB 21; Length 24;
OY      Best Local Similarity 100.0%; Pred. No. 0.16;
OY      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB      1089 TGATGAGTTGGGGCCTCAATAT 1112
DB      24 TGATGAGTTGGGGCCTCAATAT 1

RESULT 17
AAV41320
ID      AAV41320 standard; DNA: 4020 BP.
XX
XX
AC      AAV41320;
XX
XX
DT      09-NOV-1998 (first entry)
XX

```

```

DE      Human angiotensin converting enzyme gene.
XX
XX      Angiotensin converting enzyme; ACE; hypertension; exercise; human;
KM      genetic marker; ss.
XX
XX      Homo sapiens.
XX
XX      Key Location/Qualifiers
XX      CDS 23..3943
XX      FT /*tag= a
XX      FT sig-peptide 23..109
XX      FT /*tag= b
XX      FT mat-peptide 110..3940
XX      FT /*tag= c
XX
XX      WO9831835-A1.
XX
XX      23-JUL-1998.
XX
XX      22-DEC-1997; 97WO-US22974.
XX
XX      27-MAY-1997; 97US-0048309.
XX      16-JAN-1997; 97US-0035382.
XX
XX      (UYMA-) UNIV MARYLAND BALTIMORE.
XX      (UYPI-) UNIV PITTSBURGH.
XX
XX      Ferrell RE, Hagberg JM;
XX
XX      WPI; 1998-414128/35.
XX      P-PDB; AAM68155.
XX
XX      Analysis of genetic markers to identify subjects who will benefit
PT      from exercise - also assessing risk of cardiovascular disease from
PT      angiotensin-converting enzyme genotype
XX
XX      Disclosure: Page 35-41; 61pp; English.
XX
XX      This is the human angiotensin converting enzyme (ACE) gene. The
CC      gene is polymorphic with 2 common alleles (I and D), resulting in
CC      3 genotypes, II, ID and DD. It is an object of the invention to
CC      identify individuals possessing a certain genotype and associated
CC      ailment, and to determine if the health of that individual can be
CC      improved by altering behavior. A claimed method comprises
CC      identifying individuals having a certain phenotype, determining
CC      the presence or absence of genetic markers associated with the
CC      phenotype, and instituting a lifestyle change to exploit or
CC      counteract the phenotype expressed by the gene marker. If the
CC      phenotype is hypertension, the gene marker is at least one
CC      insertion (I) ACE allele and exercise training is instituted to
CC      decrease systolic and diastolic blood pressure. The gene marker
CC      can be identified by PCR amplification (see AAV41321-22) of the
CC      appropriate gene region. The general method can be used to
CC      identify subjects who will benefit most from physical exercise
CC      and also to identify those who are likely to be successful in
CC      sports.
XX
XX      Sequence 4020 BP; 860 A; 1265 C; 1169 G; 726 T; 0 other;
SQ
OY      Query Match 1.18; Score 22; DB 19; Length 4020;
OY      Best Local Similarity 100.0%; Pred. No. 1.6;
OY      Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB      922 CAGAGGCCAAGAGCTGTATG 943
DB      317 CAGAGGCCAAGAGCTGTATG 338

RESULT 18
AAA38330
ID      AAA38330 standard; DNA: 4020 BP.
XX
XX
AC      AAA38330;
XX

```

XX 21-AUG-2000 (first entry)  
 DT Human angiotensin-converting enzyme (ACE) coding region.  
 DE  
 XX Angiotensin-converting enzyme gene; ACE; coding region; polymorphism;  
 KW polymorphic marker; cardiovascular disease; myocardial infarction;  
 KM unstable angina; hypertension; atherosclerosis; stroke; prognosis;  
 XX drug screening; treatment outcome; human; ds.  
 OS Homo sapiens.  
 XX WO200022166-A2.  
 XX 20-APR-2000.  
 XX 13-OCT-1999; 99WO-IB01678.  
 XX 14-OCT-1998; 98US-0104286.  
 PR 14-OCT-1998; 98US-0104302.  
 XX (EURO-) EURONA MEDICAL AB.  
 PA Norberg LT, Andersson MK, Lindstrom PHR, Jonsson L;  
 XX WPI: 2000-318010/27.  
 DR  
 XX Assessing cardiovascular status in humans involves comparing test  
 PT polymorphic pattern comprising polymorphic positions within genes  
 PT encoding specific proteins, with reference polymorphic pattern  
 PS  
 XX Disclosure: Page 114-115; 126pp; English.

XX The invention relates to a novel method of assessing the cardiovascular  
 CC status in an individual and to newly identified polymorphisms in the  
 CC genes encoding angiotensin-converting enzyme (ACE), angiotensin II  
 CC receptor type 1 (AT1) and type 2 (AT2), angiotensinogen (AGT), renin,  
 CC aldosterone synthase, endothelin receptor type A and beta-adrenergic  
 CC receptors 1 and 2. The method comprises determining the sequence at one  
 CC or more polymorphic positions within these genes, and comparing the  
 CC pattern of polymorphisms from the individual with a reference polymorphic  
 CC pattern obtained from a population of individuals exhibiting a  
 CC predetermined cardiovascular disease status. The polymorphic markers are  
 CC useful for determining the predisposition of an individual to  
 CC cardiovascular disorders such as myocardial infarction, unstable angina,  
 CC hypertension, atherosclerosis and stroke. They are also useful for  
 CC predicting the likely cardiovascular status of a patient given a  
 CC treatment regimen comprising administration of cardiovascular drugs  
 CC (e.g., ACE inhibitors, beta-adrenergic receptor antagonists (beta-  
 CC blockers) or calcium channel blockers). One or more polymorphic markers  
 CC provides a basis for predicting the outcome of a treatment regimen.  
 CC Fragments of the genes comprising a polymorphic site may be used as  
 CC primers and probes for detecting genetic polymorphisms or in molecular  
 CC library arrays for high throughput screening. The genes, and the proteins  
 CC they encode are useful in the screening of potential cardiovascular  
 CC drugs. Determination of an individual's polymorphic pattern reduces or  
 CC eliminates trial and error in selecting a treatment for a particular  
 CC individual cardiovascular patient. It also provides the ability to  
 CC eliminate patients from clinical trials who are predicted to be  
 CC non-responsive, or at a risk for an adverse response, to a particular  
 CC treatment regimen. Adverse results in an early trial can be evaluated to  
 CC identify polymorphic patterns so that the adverse results can be  
 CC correlated with a sub-population of the test population, permitting  
 CC exclusion of such sub-populations from the treatment group. Beneficial  
 CC drugs can be approved for use in the appropriate population, thereby  
 CC decreasing the number of patients required for a clinical trial, which in  
 CC turn decreases the duration and cost of such trials. Sequences A38328 and  
 CC A38330 represent, respectively, intron 16 and the coding region of  
 CC the human ACE gene (GenBank X62855, J04144). The polymorphic sites  
 CC identified are 375A/C, 582C/T, 731A/G, 1060G/A, 1215C/T, 2193G/A,  
 CC 2228A/G, 2741G/T, 3132C/T, 3387T/C, 3503G/C, 3906G/A; and a deletion of  
 CC nucleotides 1451-1783 in intron 16.  
 CC  
 XX

SQ Sequence 4020 BP; 857 A; 1261 C; 1174 G; 728 T; 0 other;  
 Query Match 1.1%; Score 22; DB 21; Length 4020;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 922 CAGAGGCCAAGAGCTGTATG 943  
 Db 317 CAGAGGCCAAGAGCTGTATG 338

RESULT 19  
 AAK9395  
 ID AAK9395 standard; DNA: 4022 BP.  
 XX  
 AC AAK9395;  
 XX  
 DT 27-JUN-2002 (first entry)  
 XX  
 DE DNA of APP related human homologue hCP51674.  
 XX  
 KW Neuroprotective; nootropic; transgenic fly; Alzheimer's disease; Abeta;  
 KW amyloid precursor protein; tissue-specific expression control; human APP;  
 KW APP pathway modulator; gene therapy; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 148..3945  
 FT /\*tag= a  
 FT /product= "Protein of human homologue hCP51674"  
 FT /note= "No start codon"  
 XX  
 PN WO200226820-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 01-OCT-2001; 2001WO-EP11345.  
 XX  
 PP 29-SEP-2000; 2000US-236893P.  
 PR 14-JUN-2001; 2001US-298309P.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 XX  
 PI Cohen D, Dengler UJ, Finelli AL, Freuler F, Konsolaki M;  
 PI Reinhardt MMH, Zisman S;  
 XX  
 DR WPI: 2002-315796/35.  
 DR P-PSDB; AAO20501.  
 XX  
 PT New transgenic fly, containing DNA encoding an Abeta portion of human  
 PT APP, useful for identifying agents which modulate the APP pathway and  
 PT which can be used to treat Alzheimer's disease -  
 PS  
 XX Example 4; Page 93-94; 129pp; English.

XX The invention relates to a transgenic fly whose genome comprises DNA  
 CC encoding a polypeptide having the Abeta portion of human amyloid  
 CC precursor protein (APP), fused to a signal sequence. The DNA sequence  
 CC encodes a 123 (Abeta40) or 129 (Abeta42) amino acid sequence, given in  
 CC the specification. The DNA sequence is operably linked to a tissue-  
 CC specific expression control sequence. Expression of the sequence gives  
 CC the fly an altered phenotype. The purpose of the invention is for  
 CC identifying agents that inhibit or promote the expression and/or function  
 CC of genes or encoded polypeptides which modify the APP pathway. The agent  
 CC is a compound, triple helix DNA, antisense oligonucleotide, double  
 CC stranded RNA molecule, ribozyme, or particularly an antibody. It is used  
 CC to treat conditions such as Alzheimer's disease. The agent can be used as  
 CC an APP pathway modulator or in gene therapy. This polynucleotide sequence  
 CC represents the DNA of the APP related human homologue hCP51674.  
 CC  
 XX Sequence 4022 BP; 856 A; 1262 C; 1174 G; 728 T; 2 other;  
 XX

Query Match 1.1%; Score 22; DB 24; Length 4022;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943  
DB 319 CAGAGGCCAAGAGCTGTATG 340

RESULT 20  
AA004027  
ID AA004027 standard; CDNA; 4024 BP.  
XX  
AC AA004027;  
XX  
DT 07-SEP-1990 (first entry)  
XX  
DE Human angiotensin converting enzyme (ACE) gene.  
XX  
KW human angiotensin converting enzyme; hypertension;  
KW bradykinin;  
XX  
OS synthetic.  
XX  
FH Key Location/Qualifiers  
FT 23..51  
FT sig\_peptide /tag= a  
FT CDS 23..3944  
FT /\*tag= b  
XX  
PN WO9003435-A.  
XX  
PD 05-APR-1990.  
XX  
PE 29-SEP-1989; 89WO-FR00469.  
XX  
PR 27-SEP-1988; 88FR-0012620.  
XX  
PA (INRM ) INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE.  
XX  
PI Soubrier F, Alhenc-Gelas F, Hubert C, Corvol P;  
XX  
DR WPI: 1990-058128/08.  
DR P-PSDB; AAR04111.  
XX  
PT New DNA encoding human angiotensin converting enzyme used eg in  
PT diagnosis of hypertension, evaluation of enzyme inhibitors  
XX  
PS Disclosure; ; P; French.  
XX  
CC This sequence encodes pre-ACE. In the mature protein the 29 amino acid  
CC signal peptide is absent. ACE hydrolyses angiotensin I and/or kinins,  
CC most notably bradykinin. Vectors contg. the given sequence are used to  
CC produce ACE for therapeutic use eg to hydrolyse kinins implicated in  
CC inflammation.  
XX  
SQ Sequence 4024 BP; 857 A; 1263 C; 1175 G; 729 T; 0 other;

Query Match 1.1%; Score 22; DB 11; Length 4024;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943  
DB 317 CAGAGGCCAAGAGCTGTATG 338

RESULT 21  
AA035850  
ID AA035850 standard; DNA; 4024 BP.  
XX  
AC AA035850;

XX  
DT 14-JUL-1999 (first entry)  
XX  
DE Human angiotensin converting enzyme DNA.  
XX  
KW Antisense oligonucleotide; mammalian; angiotensin converting enzyme;  
KW ACE; hypertension; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9915643-A2.  
XX  
PD 01-APR-1999.  
XX  
PF 25-SEP-1998; 98WO-US20121.  
XX  
PR 25-SEP-1997; 97US-0059661.  
XX  
PA (UYFL ) UNIV FLORIDA.  
XX  
PI Mohuczy D, Phillips MI;  
XX  
DR WPI: 1999-276982/23.  
XX  
PT New antisense oligonucleotide compositions  
XX  
PS Disclosure; Page 142-143; 145pp; English.  
XX  
CC The present sequence represents human angiotensin converting enzyme  
CC (ACE) DNA. The specification describes antisense oligonucleotides  
CC (AA035853-49) for mammalian angiotensin converting enzyme (ACE) mRNA.  
CC The antisense oligonucleotides can be used for treating or preventing  
CC hypertension. The antisense oligonucleotides can also be  
CC labeled and used for visualising ACE mRNA in a cell, and for  
CC producing transgenic animals.  
XX  
SQ Sequence 4024 BP; 857 A; 1263 C; 1175 G; 729 T; 0 other;

Query Match 1.1%; Score 22; DB 20; Length 4024;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAGGCCAAGAGCTGTATG 943  
DB 317 CAGAGGCCAAGAGCTGTATG 338

RESULT 22  
AAH57430  
ID AAH57430 standard; CDNA; 5005 BP.  
XX  
AC AAH57430;  
XX  
DT 10-SEP-2001 (first entry)  
XX  
DE Human intestine cell specific CDNA sequence SEQ ID NO: 270.  
XX  
KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle;  
KW lung; liver; uterus; ovary; stomach; kidney; pancreas; ss;  
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;  
KW neuroprotective; gene therapy; cancer; Immunopathology; neuropathology.  
XX  
OS Homo sapiens.  
XX  
PN WO200132927-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 02-NOV-2000; 2000WO-US30396.  
XX  
PR 04-NOV-1999; 99US-0163508.  
XX  
PA (INCY-) INCYTE GENOMICS INC.



XX AAK68974;  
AC  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:23786.  
XX  
KW Human; Immune; hematopoietic; Immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
PF  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226886.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.



```

PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 23786; 3071bp + Sequence Listing; English.
XX
CC AA54951 to AA64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AA62170 to AA61921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AA64703
CC to AA87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AA54942 to AA54950 and AA62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 5681 BP; 1206 A; 1760 C; 1554 G; 1161 T; 0 other;

Query Match 1.0%; Score 20; DB 22; Length 5681;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1575 CAACATGAGCAGTGTCCCG 1594
   |||||||||||||||||||||
Db 3580 CAACATGAGCAGTGTCCCG 3561

Search completed: July 5, 2003, 13:28:47
Job time : 468 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## SUMMARIES

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 11:30:39 ; Search time 539f Seconds

(without alignments)  
11061.334 Million cell updates/sec

Title: US-09-937-521-15

Sequence: 1 ttcacgtgctactactatg.....ttgaagtgtcactactag 2049

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 20

Total number of hits satisfying chosen parameters: 119

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Genembl:\*

1: gb\_ba:\*  
2: gb\_bhg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pl:\*  
10: gb\_pl:\*  
11: gb\_pl:\*  
12: gb\_pl:\*  
13: gb\_pl:\*  
14: gb\_pl:\*  
15: gb\_pl:\*  
16: gb\_pl:\*  
17: gb\_pl:\*  
18: gb\_pl:\*  
19: gb\_pl:\*  
20: gb\_pl:\*  
21: gb\_pl:\*  
22: gb\_pl:\*  
23: gb\_pl:\*  
24: gb\_pl:\*  
25: gb\_pl:\*  
26: gb\_pl:\*  
27: gb\_pl:\*  
28: gb\_pl:\*  
29: gb\_pl:\*  
30: gb\_pl:\*  
31: gb\_pl:\*  
32: gb\_pl:\*  
33: gb\_pl:\*  
34: gb\_pl:\*  
35: gb\_pl:\*  
36: gb\_pl:\*  
37: gb\_pl:\*  
38: gb\_pl:\*  
39: gb\_pl:\*  
40: gb\_pl:\*  
41: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	100.0	3108	10	AB037111	AB037111 Mus muscu
2	2016	98.4	4835	10	AB037181	AB037181 Mus muscu
3	1493	72.9	1571	10	BC022604	BC022604 Mus muscu
4	192	9.4	165407	2	AC117816	AC117816 Mus muscu
5	179	8.7	165407	2	AC117816	AC117816 Mus muscu
6	179	8.7	184101	2	AC113485	AC113485 Mus muscu
7	72	3.5	184101	2	AC113485	AC113485 Mus muscu
8	68	3.3	2283	6	E50433	E50433 Novel prote
9	68	3.3	2456	10	AB057433	AB057433 Rattus no
10	53	2.6	539	6	E50435	E50435 Novel prote
11	47	2.3	235	9	HUMDAC12M5	AF250847 Homo sapi
12	47	2.3	2289	9	AF250847	AF250847 Homo sapi
13	47	2.3	2507	6	AX207121	AX207121 Sequence
14	47	2.3	162228	2	AL450382	AL450382 Human DNA
15	47	2.3	176865	2	AL513306	AL513306 Homo sapi
16	47	2.3	178899	2	AL589794	AL589794 Human DNA
17	39	1.9	110000	2	AC055726_2	Continuation (3 of
18	36	1.8	105838	2	AC105714	AC105714 Rattus no
19	32	1.6	105838	2	AC105714	AC105714 Rattus no
20	30	1.5	187130	2	AL583858	AL583858 Homo sapi
21	25	1.2	38	6	AX207126	AX207126 Sequence
22	23	1.1	164805	8	OSJN00174	AL662970 Oryza sat
23	23	1.1	175330	2	AC012131	AC012131 Homo sapi
24	22	1.1	168	9	FI93462S02	FI93463 Pan trogl
25	22	1.1	4020	6	ARI37383	ARI37383 Sequence
26	22	1.1	4020	6	HUMAICEB	J04144 Human angio
27	22	1.1	4022	6	AX429555	AX429555 Sequence
28	22	1.1	4024	6	A00914	A00914 H. sapiens g
29	22	1.1	24070	6	AF118569	AF118569 Homo sapi
30	22	1.1	80858	9	AC007360	AC007360 Homo sapi
31	22	1.1	201500	2	AC103946	AC103946 Mus muscu
32	21	1.0	21	6	E50434	E50434 Novel prote
33	21	1.0	4014	10	AF201331	AF201331 Rattus no
34	21	1.0	4014	10	AF201332	AF201332 Rattus no
35	21	1.0	4014	10	RND03708	U03708 Rattus norv
36	21	1.0	4142	10	RND03734	U03734 Rattus norv
37	21	1.0	53423	9	AL591419	AL591419 Human DNA
38	21	1.0	76700	2	AC131469	AC131469 Rattus no
39	21	1.0	100858	2	AC096435	AC096435 Rattus no
40	21	1.0	179743	2	AC098290	AC098290 Rattus no
41	21	1.0	181663	2	AC068982	AC068982 Homo sapi
42	21	1.0	186651	2	AC120772	AC120772 Rattus no
43	21	1.0	191192	2	AC106294	AC106294 Rattus no
44	21	1.0	197174	2	AC094442	AC094442 Rattus no
45	21	1.0	212255	2	AC096705	AC096705 Rattus no
46	20	1.0	750	11	AL592564	AL592564 Meleagris
47	20	1.0	1065	5	HSMB00587	AL080086 Homo sapi
48	20	1.0	1680	5	AF231711	AF231711 Xenopus l
49	20	1.0	17411	1	MYCRN1	L35043 Mycoplasma
50	20	1.0	23461	2	AC116980	AC116980 Dictyoste
51	20	1.0	43928	2	AC090263	AC090263 Homo sapi
52	20	1.0	53772	2	AC097284	AC097284 Rattus no
53	20	1.0	60657	2	AC101392	AC101392 Mus muscu
54	20	1.0	71669	2	AC102279	AC102279 Mus muscu
55	20	1.0	90820	2	AC111155	AC111155 Homo sapi
56	20	1.0	97143	9	AC025756	AC025756 Homo sapi
57	20	1.0	97488	8	ATF29H11	AL049659 Arabidops
58	20	1.0	110000	10	AE014173_1	Continuation (2 of
59	20	1.0	118873	9	HS1184F4	AL034450 Human DNA
60	20	1.0	141418	9	AC011741	AC011741 Homo sapi
61	20	1.0	142500	2	AC117145	AC117145 Rattus no
62	20	1.0	144091	8	AP003022	AP003022 Oryza sat
63	20	1.0	144269	2	AC098342	AC098342 Rattus no
64	20	1.0	144734	2	AC123948	AC123948 Mus muscu
65	20	1.0	145945	2	AC119369	AC119369 Rattus no

C	66	20	1.0	1.148526	10	AL732503	AL732503	Mouse	DNA
C	67	20	1.0	1.149768	2	AC016239	AC016239	Homo sapi	
C	68	20	1.0	1.150683	2	AC120517	AC120517	pan trogl	
C	69	20	1.0	1.151371	2	AC091889	AC091889	Homo sapi	
C	70	20	1.0	1.154373	10	AC121988	AC121988	Mus muscu	
C	71	20	1.0	1.157821	2	AC120725	AC120725	Rattus no	
C	72	20	1.0	1.159283	2	AC107705	AC107705	Mus muscu	
C	73	20	1.0	1.161739	9	AC105277	AC105277	Homo sapi	
C	74	20	1.0	1.163040	2	AC112470	AC112470	Rattus no	
C	75	20	1.0	1.163312	9	AC007955	AC007955	Homo sapi	
C	76	20	1.0	1.165614	2	AC125707	AC125707	Rattus no	
C	77	20	1.0	1.167340	2	AC111314	AC111314	Rattus no	
C	78	20	1.0	1.167658	2	AC123942	AC123942	Mus muscu	
C	79	20	1.0	1.168440	2	AC109112	AC109112	Rattus no	
C	80	20	1.0	1.170325	2	AC109013	AC109013	Rattus no	
C	81	20	1.0	1.173840	9	AC092445	AC092445	Homo sapi	
C	82	20	1.0	1.174711	2	AC025515	AC025515	Homo sapi	
C	83	20	1.0	1.175080	2	AC113545	AC113545	Rattus no	
C	84	20	1.0	1.176951	2	AL845340	AL845340	Mus muscu	
C	85	20	1.0	1.177172	9	CNS00006	CNS00006	Human chr	
C	86	20	1.0	1.178298	2	AC094759	AC094759	Rattus no	
C	87	20	1.0	1.178652	10	AL713853	AL713853	Mouse DNA	
C	88	20	1.0	1.180682	2	AC109776	AC109776	Rattus no	
C	89	20	1.0	1.182199	2	AC095097	AC095097	Rattus no	
C	90	20	1.0	1.185306	2	AC013570	AC013570	Homo sapi	
C	91	20	1.0	1.185761	2	AC108814	AC108814	Mus muscu	
C	92	20	1.0	1.186314	10	AC003654	AC003654	Mus muscu	
C	93	20	1.0	1.187507	2	AC124724	AC124724	Mus muscu	
C	94	20	1.0	1.188061	2	AC007373	AC007373	Homo sapi	
C	95	20	1.0	1.188561	10	AL591905	AL591905	Mouse DNA	
C	96	20	1.0	1.188921	9	AC096730	AC096730	Homo sapi	
C	97	20	1.0	1.189891	2	AC119058	AC119058	Papio cyn	
C	98	20	1.0	1.192802	2	AC103600	AC103600	Mus muscu	
C	99	20	1.0	1.193865	2	AC122270	AC122270	Mus muscu	
C	100	20	1.0	1.196664	2	AC131185	AC131185	Mus muscu	
C	101	20	1.0	1.198554	2	AC090552	AC090552	Homo sapi	
C	102	20	1.0	1.200065	10	AC080021	AC080021	Genomic s	
C	103	20	1.0	1.201839	2	AC098333	AC098333	Rattus no	
C	104	20	1.0	1.202066	2	AC110633	AC110633	Rattus no	
C	105	20	1.0	1.202249	2	AC095427	AC095427	Rattus no	
C	106	20	1.0	1.206819	2	AC016029	AC016029	Homo sapi	
C	107	20	1.0	1.207538	2	AC106886	AC106886	Homo sapi	
C	108	20	1.0	1.209171	2	AL845279	AL845279	Mus muscu	
C	109	20	1.0	1.210321	2	AC114001	AC114001	Mus muscu	
C	110	20	1.0	1.213473	2	AC094371	AC094371	Rattus no	
C	111	20	1.0	1.213901	2	AC093472	AC093472	Mus muscu	
C	112	20	1.0	1.216238	10	AL513356	AL513356	Mouse DNA	
C	113	20	1.0	1.218553	2	AC079940	AC079940	Mus muscu	
C	114	20	1.0	1.221712	2	AC126447	AC126447	Mus muscu	
C	115	20	1.0	1.223368	2	AC103652	AC103652	Mus muscu	
C	116	20	1.0	1.235426	2	AC121074	AC121074	Canis fam	
C	117	20	1.0	1.240233	2	AC127342	AC127342	Mus muscu	
C	118	20	1.0	1.258661	2	AC129078	AC129078	Mus muscu	
C	119	20	1.0	1.269423	2	AC129333	AC129333	Mus muscu	

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AB037111										
AB037111	3108 bp	mRNA	linear							
		Mus musculus	LCDease mRNA for neutral ceramidase, complete cds.							
AB037111										
AB037111.1	GI:7576308	neutral ceramidase.								
		Mus musculus liver cDNA to mRNA.								
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (sites)								
		Tani,M., Okino,N., Mori,K., Tanigawa,T., Ito,H. and Ito,M.								
		Molecular cloning of the full-length cDNA encoding mouse neutral								

JOURNAL.	neutral/alkaline ceramidases
MEDLINE	J. Biol. Chem. 275 (15), 11229-11234 (2000)
REFERENCE	20219171
AUTHORS	2 (bases 1 to 3108)
TITLE	Okino, N. and Ito, M.
JOURNAL	Direct Submission
	Submitted (14-JUN-2000) Makoto Ito, Kyushu University, Graduate School of Bioresource and Bioenvironmental Sciences, Department of Bioscience and Biotechnology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan (E-mail:makoto@agr.kyushu-u.ac.jp, Tel:+81-92-642-2900, Fax:+81-92-642-2900)
FEATURES	
Source	Location/Qualifiers
	1..3108
gene	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/tissue_type="liver"
CDS	1..3108
	/gene="Lcdase"
	726..2996
	/gene="Lcdase"
	/codon_start=1
	/product="neutral ceramidase"
	/protein_id="BAAG4545.1"
	/db_xref="GI:7576309"
	/translation="MAKRTFSTLEAFLLFLVITVTVALLLTLFTVSGTIEHKKDSD GNMFSTLSSTTPPTPPITQTPNPFSPFRNSGYIGRADCDGVSDILMKGKKN GONARGLIHLFSRAFLIADPDGSNNRFAVESELCMSORLEVLRLRESYSLFR RDNYLSAIHTHSFGPFOYTYLILASEGSMNTFOYIVSGIMKSIDIAHTNLPGMG IFIKGVANVOINRSPSYLYLNPSREARSNTDKMLVKLVLDNGEDGLISVPNIIL AIIHVSMNNSNHFPVSNMYCAALTFDEKRNKYLPDGPVAGFASNLDGSDPNILII GPCVNTGESCDNDKSTCPNGSPSCMASGCGDMFEETHIIGTIITYOKAKELAYASAS OEYVGPIVAHQVMNMVDVSVOLNATHRVTKCKPALGSEFAGTIDVSGINTIQTGT EGCFEMDTLRDLQKRPSEEIVECOEKPRPLIHSGELTIPHPMDPIDVDOIVTVGSS AIAAPIPELTMSGRREREAIRKEFALYGMDMTVGYLAGLSNVTHYTYYEEYOAOBR YEASTIYEPHTLSAVIOLEFDLAKALATDVAMSSGSEPPEFKNLIASLIPMIADB APICKHGVDVOPAKPERKRYENGEVEVFYCANPKNSMENOHOFLEYEKEDSVAPR AIMXNDSWETRFYWKGIJGLSNATTIWMHLPDAITPFIIRIFIRIGHNRKDELKPAAV ILAEGISSPEEVTT"
BASE COUNT	888 a 760 c 714 g 746 t
ORIGIN	
Query Match	100.0%; Score 2049; DB 10; Length 3108;
Best Local Similarity	100.0%; Prid. No. 0;
Matches 2049;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 TTTCAGTGCTCATCATTGGCGTTGGAGACGGATTGCACAGACAAGTCTCAGATATC 60
Dd	948 TTTCAGTGCTCATCATTGGCGTTGGAGACGGATTGCACAGACAAGTCTCAGATATC 10077
OY	61 AATTGATGGCTATGGCAAAAATGGCCAGAATCACGGGGTCTCTCCACGAGCTGTT 120
Dd	1008 AATTGATGGCTATGGCAAAAATGGCCAGAATCACGGGGTCTCTCCACGAGCTGTT 10677
OY	121 AGCCGTGCTTATCTTGGCGGATCCAGATGGGTCGAATGCAATGGCATTTGTGAGCTG 180
Dd	1068 AGCCGTGCTTATCTTGGCGGATCCAGATGGGTCGAATGCAATGGCATTTGTGAGCTG 11227
OY	181 GAATATGATGATTTGCCAACGACTGAGTTGGAGGTCCTGAAGAGACTAGAGATTA 240
Dd	1128 GAATATGATGATTTGCCAACGACTGAGTTGGAGGTCCTGAAGAGACTAGAGATTA 11877
OY	241 TATGGCTCTCTGATTCGAGAGACAATGTTATCTCTGAGTGGCATTTCCACACACTCTGGC 300
Dd	1188 TATGGCTCTCTGATTCGAGAGACAATGTTATCTCTGAGTGGCATTTCCACACACTCTGGC 12477
OY	301 CCACGAGGGTTTTCCAAATATACACTCTATATCTCGCCACGAGGAGATTGACGACC 360
Dd	1248 CCACGAGGGTTTTCCAAATATACACTCTATATCTCGCCACGAGGAGATTGACGACC 13077
OY	361 ACCTTTAGTATAGTCTCTGGGATATGAAGAGCATGATATATAGTCTCACACAATCTT 420
Dd	1308 ACCTTTAGTATAGTCTCTGGGATATGAAGAGCATGATATATAGTCTCACACAATCTT 13677

OY	421	AAACGAGCAAAATCTTTATACAAAGGAATGTGTATGTGCATCAACCGAAC	480
Db	1368	AAACGAGCAAAATCTTTATACAAAGGAATGTGTATGTGCATCAACCGAAC	1427
OY	481	CCCTCTCTTACCTTCTGTAATCCACAGTCAGAGAGAGAAAGTATCTTCAACACAGAC	540
Db	1428	CCCTCTCTTACCTTCTGTAATCCACAGTCAGAGAGAGAAAGTATCTTCAACACAGAC	1487
OY	541	AAGAAATGCTGTGCTTGTGAACCTGTGATTTGTAATGAGAAAGACTTGGCTTATCAGC	600
Db	1488	AAGAAATGCTGTGCTTGTGAACCTGTGATTTGTAATGAGAAAGACTTGGCTTATCAGC	1547
OY	601	TGTTTGGCAATCCACCCCGTGAGATGAACAATAGCAACCACTTTGGTAATAGACAAT	660
Db	1548	TGTTTGGCAATCCACCCCGTGAGATGAACAATAGCAACCACTTTGGTAATAGACAAT	1607
OY	661	ATGGGCTATGCGGCTTACCTTTTGGACACAAAGAAACAAAGCTATATCTCGTGGACAG	720
Db	1608	ATGGGCTATGCGGCTTACCTTTTGGACACAAAGAAACAAAGCTATATCTCGTGGACAG	1667
OY	721	GGACCGTTTGTAGACGCTTGTCTTCAATCAAATCTCGAGACGTGTCAACCCACATTTCTT	780
Db	1668	GGACCGTTTGTAGACGCTTGTCTTCAATCAAATCTCGAGACGTGTCAACCCACATTTCTT	1727
OY	781	GGCCCGCATGTGTGCACACACAGGGGAGTCTGTGACAAACGACAGAGCAGCTGTCCCAAC	840
Db	1728	GGCCCGCATGTGTGCACACACAGGGGAGTCTGTGACAAACGACAGAGCAGCTGTCCCAAC	1787
OY	841	GGTGGGCTTAGCATGTGCATGGCCAGCGGACCTTGACAAAGCATGTTTGAAGACACAC	900
Db	1788	GGTGGGCTTAGCATGTGCATGGCCAGCGGACCTTGACAAAGCATGTTTGAAGACACAC	1847
OY	901	ATTATAGAGCGATCATCTATCAGAAAGCCAAAGAGCTGTATTCCTTGCCCTCCAGAG	960
Db	1848	ATTATAGAGCGATCATCTATCAGAAAGCCAAAGAGCTGTATTCCTTGCCCTCCAGAG	1907
OY	961	GTGACCGGCCAGTGTGGAGCTCACACAGTGGGTGAACATGACAGATGTGAGGCTCCAG	1020
Db	1908	GTGACCGGCCAGTGTGGAGCTCACACAGTGGGTGAACATGACAGATGTGAGGCTCCAG	1967
OY	1021	CTCAATGCCACACACACAGTGAAAGCGTGTAAACCTGCCCTGGGCTACAGTTTGGCGCA	1080
Db	1968	CTCAATGCCACACACACAGTGAAAGCGTGTAAACCTGCCCTGGGCTACAGTTTGGCGCA	2027
OY	1081	GGCAATTTGATGAGATTTGGGGCTCAATATTACACAGGGACTACGGAAGGGATCCA	1140
Db	2028	GGCAATTTGATGAGATTTGGGGCTCAATATTACACAGGGACTACGGAAGGGATCCA	2087
OY	1141	TTTCGGGACACTCTTCGGGACCACTCTGTGGGAAACCATCTGGAAGATTTGTAGAGTG	1200
Db	2088	TTTCGGGACACTCTTCGGGACCACTCTGTGGGAAACCATCTGGAAGATTTGTAGAGTG	2147
OY	1201	CAGAAACCCAACCAATCTGCTTACAGTGTGAGAGCTGAGGATACCAATCTTTGGCAA	1260
Db	2148	CAGAAACCCAACCAATCTGCTTACAGTGTGAGAGCTGAGGATACCAATCTTTGGCAA	2207
OY	1261	CCAATATTTGTGATGTTTCAGATTTTACGTTGGGTCTTGGCCATATGCTTATCCCT	1320
Db	2208	CCAATATTTGTGATGTTTCAGATTTTACGTTGGGTCTTGGCCATATGCTTATCCCT	2267
OY	1321	GGGGAATTTAACCAACCATGTGGGGACGAAGATTTGTGTAAGGTAATTAAGAAATTTGCA	1380
Db	2268	GGGGAATTTAACCAACCATGTGGGGACGAAGATTTGTGTAAGGTAATTAAGAAATTTGCA	2327
OY	1381	CTTTATGGAGTGAAGATATGACCGTTTGTATTCGACAGTCTAAGCAATGTTTATACAT	1440
Db	2328	CTTTATGGAGTGAAGATATGACCGTTTGTATTCGACAGTCTAAGCAATGTTTATACAT	2387
OY	1441	TACATTCACATATGAAAGATATCCAGGTCACAGGATACAGAGATCTACAAATCTAT	1500
Db	2388	TACATTCACATATGAAAGATATCCAGGTCACAGGATACAGAGATCTACAAATCTAT	2447

QY	1501	GGACCAACAACCCGTGTGGATACATCCAACTCTTGGAGACCTTGGTAAGGCATATGCT	1560
Db	2448	GGACCAACAACCCGTGTGTGCATACATCCAACTCTTGGAGACCTTGGTAAGGCATATGCT	2507
QY	1561	ACGGACACAGTAGCCCAACATGAGCAGTGGTCCGAGCCTCCATTTCTTCAAAAATCTATA	1620
Db	2508	ACGGACACAGTAGCCCAACATGAGCAGTGGTCCGAGCCTCCATTTCTTCAAAAATCTATA	2567
QY	1621	GCTTCACTTATTTCTTAATATTGGGGATAGAGACACCAATTGGCAAACTTTTGGGGATGTC	1680
Db	2568	GCTTCACTTATTTCTTAATATTGGGGATAGAGACACCAATTGGCAAACTTTTGGGGATGTC	2627
QY	1681	TTTGAGGCACAAAACCTGAAATACAGTGGGGGAAAGTGGTTGAGATTAATTGTATGGC	1740
Db	2628	TTTGAGGCACAAAACCTGAAATACAGTGGGGGAAAGTGGTTGAGATTAATTGTATGGC	2687
QY	1741	GCTAACCCAAAGAATTACAGAGAGAACACACCATCAAACTTCTCCTACTGTGGAGAAA	1800
Db	2688	GCTAACCCAAAGAATTACAGAGAGAACACACCATCAAACTTCTCCTACTGTGGAGAAA	2747
QY	1801	TACGAGGACCTCTGAGCTGACTGGCAGATTAATGTATTAAAGATGCTCTGGGAGCAGG	1860
Db	2748	TACGAGGACCTCTGAGCTGACTGGCAGATTAATGTATTAAAGATGCTCTGGGAGCAGG	2807
QY	1861	TTTTATTGGCAAAAGAAATACATGCGTGTGAGCATGCAACATATATCTGGCATATTCCA	1920
Db	2808	TTTTATTGGCAAAAGAAATACATGCGTGTGAGCATGCAACATATATCTGGCATATTCCA	2867
QY	1921	GATACCTGCTTACCCCTGGAACTCTACAGAAATTAAGATATTTTGGACACATTCGGAAGCAGAA	1980
Db	2868	GATACCTGCTTACCCCTGGAACTCTACAGAAATTAAGATATTTTGGACACATTCGGAAGCAGAA	2927
QY	1981	CTTCTGAACCCCGCTGTCATACTAGCATTTTGAAGGAATTTCTTCTCCCTTTTGAAGTTGTC	2040
Db	2928	CTTCTGAACCCCGCTGTCATACTAGCATTTTGAAGGAATTTCTTCTCCCTTTTGAAGTTGTC	2987
QY	2041	ACTACTTAG 2049	
Db	2988	ACTACTTAG 2996	

FEATURES	source	gene
LOCUS	AB037181	
RESULT 2		
DEFINITION	AB037181	4835 bp mRNA linear ROD 08-JUN-2001
ACCESSION	AB037181	Mus musculus Ecdase mRNA for neutral ceramidase, complete cds.
VERSION	AB037181.1	GI:7576310
KEYWORDS		neutral ceramidase.
SOURCE		Mus musculus tissue_lib:brain cDNA to mRNA.
ORGANISM		Mus musculus
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathae; Muridae; Murinae; Mus.
AUTHORS		1 (sites)
TITLE		Tani, M., Okino, N., Mori, K., Tanigawa, T., Izu, H. and Ito, M.
JOURNAL		Molecular cloning of the full-length cDNA encoding mouse neutral ceramidase. A novel but highly conserved gene family of neutral/alkaline ceramidases
MEDLINE		J. Biol. Chem. 275 (15), 11229-11234 (2000)
REFERENCE		20219171
AUTHORS		2 (bases 1 to 4835)
TITLE		Okino, N. and Ito, M.
JOURNAL		Direct Submission
FEATURES		Submitted (17-JAN-2000) Makoto Ito, Kyushu University, Graduate School of Bioresource and Bioenvironmental Sciences, Department of Bioscience and Biotechnology, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan (E-mail:makoto@agr.kyushu-u.ac.jp, Tel:81-92-642-2900, Fax:81-92-642-2900)
source		location/Qualifiers
		1. 4835
		/organism="Mus musculus"
		/db_xref="taxon:10090"
		/tissue_lib="brain"
		1. 4835

```

CDS
      /gene="BCDase"
      /gene..2711
      /codon_start=1
      /product="neutral ceramidase"
      /protein_id="BA94546.1"
      /db_xref="GI:7576311"
      /translation="MAKRTSTLEAFLEFLVIMVITVALLTLFVNSGTLENKDS
GNHWFSTLGSSTTQPPPTPNPFPSFNFSGYIYGVRACTQVSDINLMGKN
GONARGLTLRLEFRAFIADPDGSRNMAFVSVELMSIQRLEVLKRESEKSLYR
RDNVLSAHTSGPAGFEQYTLTILASGFSNRTFOYIVSGISIDIAHNLKMG
IFINKGNVANOINRSPSSYLLINPSEARSNSNTDKEMLVKLDNEDJLISMF
AHPVSNNNSNHPVNSDNNGYAYLFDEKNGYLPQSGPEFAGRSASLSGYSRIL
GPHCVNTEBSCDNDSCTCPNGPSMCMASPEQDMFESTHIIIGRIIYQAKELVAS
QEVTPVLAHQVMWMTDVSYQNLNHTVTKCPALGYSFAGITDVGSLNITQGT
BEDPMDLRLDLGKPSBEIIEVCEKPKPILHSELIPHEMODIDVQIYVQSL
AIAAIPEGLTMSGRREAIKEFEALYKMDMYVIALSVYTHYITTYEYOQR
YEASTITGVPHLSAYIOLEFRLAKAIADIVANSSGSEPEPKLILSLIPNLADR
APIGHFDVLOPAKREYRVCVEVIVPGANPKNSAENQHOTPLTVEKIEDSVADM
QIMYNDASHERTFYHKKGLIGLSNMTIIMHPDIATPGIYRIRLRFHNRKQELLPAY
ITAEGISPSFEVVT"
BASE COUNT      1460 a      996 c      1048 g      1331 t
ORIGIN
Query Match      98.4% Score 2016; DB 10; Length 4835;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2016; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 34 GATTGCACAGCAGAGTGCAGATATCAATTTGATGGCTATGCAAAATGGCCAGAT 93
Db 696 GATTGCACAGCAGCAGTGCATATATCAATTTGATGGCTATGCAAAATGGCCAGAT 755
OY 94 GCACGGGGCTCTCTCACCAGGCTGTTCACGCTCTTTATCTTGCGGATCCAGATGGG 153
Db 756 GCACGGGGCTCTCTCACCAGGCTGTTCACGCTCTTTATCTTGCGGATCCAGATGGG 815
OY 154 TCAAAATCGAATGGCAATTTGTGAGCGGAGCAATGATGATTTCCCAAGGACTGAGGTTG 213
Db 816 TCAAAATCGAATGGCAATTTGTGAGCGGAGCAATGATGATTTCCCAAGGACTGAGGTTG 875
OY 214 GAGGTCTCTGAAGAGACTAGAGATAAATATGGCTCTCTGTATCGAAGAGACAATGTTATC 273
Db 876 GAGGTCTCTGAAGAGACTAGAGATAAATATGGCTCTCTGTATCGAAGAGACAATGTTATC 935
OY 274 CTGAGTGCCATTTCACACACACTCTGCCCCAGCAGGGTTTTCATATACACTCTATATA 333
Db 936 CTGAGTGCCATTTCACACACACTCTGCCCCAGCAGGGTTTTCATATACACTCTATATA 995
OY 334 CTGCGCAGGAGGGAATTCAGCAACCGGACCTTCACTAGTACATAGCTCTGCGGATCGATGAG 393
Db 996 CTGCGCAGGAGGGAATTCAGCAACCGGACCTTCACTAGTACATAGCTCTGCGGATCGATGAG 1055
OY 394 AGCATTTGATATAGCTCACACAATCTTAAACAGGCAAAATCTTTATCAACAAGGAAT 453
Db 1056 AGCATTTGATATAGCTCACACAATCTTAAACAGGCAAAATCTTTATCAACAAGGAAT 1115
OY 454 GTTGCTTAATGTCAAGTCAACCGAAGCCCTCTCTTACCTTCTGAAATCCACAGTCAGAG 513
Db 1116 GTTGCTTAATGTCAAGTCAACCGAAGCCCTCTCTTACCTTCTGAAATCCACAGTCAGAG 1175
OY 514 AGAGCAAGGTATTTCTCAACAACAGCAAGGAATGCGGTCTTGAAGTGGTGGATTTG 573
Db 1176 AGAGCAAGGTATTTCTCAACAACAGCAAGGAATGCGGTCTTGAAGTGGTGGATTTG 1235
OY 574 AATGAGAGAAGACTTGGGTCTTATCAGCTGTGTTGCATCCACCCGTGAGCATGAACAAT 633
Db 1236 AATGAGAGAAGACTTGGGTCTTATCAGCTGTGTTGCATCCACCCGTGAGCATGAACAAT 1295
OY 634 AGCAACAACCTTTGTGAATAGTACAAATATGGCTATGCGGCTTACCTTTTGAAGCAAGA 693
Db 1296 AGCAACAACCTTTGTGAATAGTACAAATATGGCTATGCGGCTTACCTTTTGAAGCAAGA 1355
OY 694 AAGAACAAGGCTATCTGCTGTGAGACAGGACCGTTGTATGACAGGCTTTGCTTCAATCAAT 753
```

```

Db 1356 AAGAACAAGGCTATCTGCTGTGACAGGACCGTTGTATGACAGGCTTTCTTCAATCAAT 1415
OY 754 CTGCGAGAGCTGTGACCAACAATCTTTGGCCCCGATTTGTGAACACAGAGGAGTCTTGT 813
Db 1416 CTGCGAGAGCTGTGACCAACAATCTTTGGCCCCGATTTGTGTGAACACAGAGGAGTCTTGT 1475
OY 814 GACAAACAGACAGACACCTGTGCCAACGGTGGGCCCTAGCATGTGCATGGCCAGCGACCT 873
Db 1476 GACAAACAGACAGACACCTGTGCCAACGGTGGGCCCTAGCATGTGCATGGCCAGCGACCT 1535
OY 874 GGACAAAGATGTTTGAAGACACACACATTTATAGACGATATCATCTATAGAAAGCCAAAG 933
Db 1536 GGACAAAGATGTTTGAAGACACACACATTTATAGACGATATCATCTATAGAAAGCCAAAG 1595
OY 934 GAGCTGTATGCTCTGCTCTCCAGAGAGGTGACCGGCCAGTCTTGCACCTCACCAAGTGG 993
Db 1596 GAGCTGTATGCTCTCTGCTCTCCAGAGAGGTGACCGGCCAGTCTTGCACCTCACCAAGTGG 1655
OY 994 GTGAACATGACAGATGTGAGCGCTCAGCTCAATGCCACACACAGTGAAGACGTGTANA 1053
Db 1656 GTGAACATGACAGATGTGAGCGCTCAGCTCAATGCCACACACAGTGAAGACGTGTANA 1715
OY 1054 CCGTCCCTGGGCTTACAGTTTTCGCGAGGCACAATTGATGAGTTTGGGGCTCAATAT 1113
Db 1716 CCGTCCCTGGGCTTACAGTTTTCGCGAGGCACAATTGATGAGTTTGGGGCTCAATAT 1775
OY 1114 ACACAGGGAAGTACGGAAGGAGATCCATTTCTGGGACACTCTTGGGACCAAGCTCTTGGGA 1173
Db 1776 ACACAGGGAAGTACGGAAGGAGATCCATTTCTGGGACACTCTTGGGACCAAGCTCTTGGGA 1835
OY 1174 AAACCATCTGAAGAGATTTAGAGTGTCAAGAACCCAAACCAATCTGCTTACAGTGA 1233
Db 1836 AAACCATCTGAAGAGATTTAGAGTGTCAAGAACCCAAACCAATCTGCTTACAGTGA 1895
OY 1234 GAGCTGACGATTCACATCTTGGCAACACAGATATTTGATGTTCAAGTTTACCGTT 1293
Db 1896 GAGCTGACGATTCACATCTTGGCAACACAGATATTTGATGTTCAAGTTTACCGTT 1955
OY 1294 GGGTCTTGGCCATACCTCTATCCCTGGGGAATTAACACAGTGTGGGAGGAAGATTT 1353
Db 1956 GGGTCTTGGCCATACCTCTATCCCTGGGGAATTAACACAGTGTGGGAGGAAGATTT 2015
OY 1354 CGTAGGCAATTTAAACAAATTTGACATTTATGGGATGAAGATGACCGTTGTATC 1413
Db 2016 CGTAGGCAATTTAAACAAATTTGACATTTATGGGATGAAGATGACCGTTGTATC 2075
OY 1414 GGAGGCTAAGCAATGTTATACACTTACATTTACACATTAAGAAGAAATACAGGCTCAG 1473
Db 2076 GGAGGCTAAGCAATGTTATACACTTACATTTACACATTAAGAAGAAATACAGGCTCAG 2135
OY 1474 CGGTACGAGGAGCATCTACAATCTATGAGACACACACCCGTCTGCATACATCCAACTC 1533
Db 2136 CGGTACGAGGAGCATCTACAATCTATGAGACACACACCCGTCTGCATACATCCAACTC 2195
OY 1534 TTTCAGAGACCTTGTCTAAGCAATTTGCTACGACACAGTATGCCAATGACAGTGGTCC 1593
Db 2196 TTTCAGAGACCTTGTCTAAGCAATTTGCTACGACACAGTATGCCAATGACAGTGGTCC 2255
OY 1594 GAGCCCTCACTCTTCAAAAATCTAATAGCTTACTTATCTTAATTTGCGATGAGACA 1653
Db 2256 GAGCCCTCACTCTTCAAAAATCTAATAGCTTACTTATCTTAATTTGCGATGAGACA 2315
OY 1654 CCAATTTGCAAACTTTTGGGATGTCTTGACAGCAGCAAAACCTGAATACAGAGTGGGA 1713
Db 2316 CCAATTTGCAAACTTTTGGGATGTCTTGACAGCAGCAAAACCTGAATACAGAGTGGGA 2375
OY 1714 GAAAGTGTGTAAGTATATTTGTAGCGCTTAACCAAGAAATTCAGCAGAGACCAAGCC 1773
Db 2376 GAAAGTGTGTAAGTATATTTGTAGCGCTTAACCAAGAAATTCAGCAGAGACCAAGCC 2435
OY 1774 CATCAAACTTCTCAGCTGTGAGAAATTCGAGGAGCTGTGTGCTGACGTGGAGATTAATG 1833
```

Db 2436 CATCAACCTTCCTCAGTGTGGAGAAATACGAGAGCTGTGACTGTGACGAGATATATG 2495  
 QY 1834 TATAAGATGCTCTGGGAGAGAGCTTTATTTGGACAAAGAAATAGTGGTCTGACC 1893  
 Db 2436 TATAAGATGCTCTGGGAGAGAGCTTTATTTGGACAAAGAAATAGTGGTCTGACC 2555  
 QY 1894 AATGCAACATATATGATGATATTCACATATCTGCTTACCCGGAATCTACAGAAATAGA 1953  
 Db 2556 AATGCAACATATATGATGATATTCACATATCTGCTTACCCGGAATCTACAGAAATAGA 2615  
 QY 1954 TATTTTGGACACATGAGAGAGAACTTGAACCCGCTGATACATAGCAATTTGAA 2013  
 Db 2616 TATTTTGGACACATGAGAGAGAACTTGAACCCGCTGATACATAGCAATTTGAA 2675  
 QY 2014 GGAATTTCTCTCTTGTGAAGTGTGCACTACTTGA 2049  
 Db 2676 GGAATTTCTCTCTTGTGAAGTGTGCACTACTTGA 2711

RESULT 3  
 BC022604  
 LOCUS 1571 bp mRNA linear ROD 07-ANG-2002  
 DEFINITION Mus musculus, Similar to N-acylsphingosine amidohydrolase 2, clone  
 IMAGE:4162513, mRNA, partial cds.  
 ACCESSION BC022604  
 VERSION BC022604.1 GI:18490446  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1571)  
 Strausberg, R.  
 Direct Submission  
 Submitted (01-FEB-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The J.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdickpaxil.stanford.edu](mailto:mdickpaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

REMARK  
 COMMENT  
 Clon distribution: MGC clone distribution information can be found  
 through the J.M.A.G.E. Consortium/LNL at: <http://image.jnl.gov>  
 Series: IRK Plate: 43 Row: C Column: 22.  
 Location/Qualifiers  
 1. 1571  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="IMAGE:4162513"  
 /tissue\_type="Liver, normal, 5 month old male mouse."  
 /clone\_id="NCI CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SORT6"  
 <1. 1493  
 /codon\_start=3  
 /product="Similar to N-acylsphingosine amidohydrolase 2"  
 /protein\_id="AAH22604.1"  
 /db\_xref="GI:18490447"  
 /translation="KLVPLNGEDJGLISMFALHPYSNNKSNHFFVSDNMGYATLFEQ  
 EKNKYLPGGQPFVAFASNSLGDVSPNLGHCYVNTGSCNDSTCCNGSPSCMA  
 SGPGDMFESTHIIIGRIIYOKAKELIYASASQVTPVLAHQVMWMTDVSVDLNAHT  
 VKTKPALGYSFAAGTIDVSGSLNTIGCTTEBDPFWDLRLDLKPSSEIVECQPK

CDS

FEATURES  
 source  
 Clon distribution: MGC clone distribution information can be found  
 through the J.M.A.G.E. Consortium/LNL at: <http://image.jnl.gov>  
 Series: IRK Plate: 43 Row: C Column: 22.  
 Location/Qualifiers  
 1. 1571  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="IMAGE:4162513"  
 /tissue\_type="Liver, normal, 5 month old male mouse."  
 /clone\_id="NCI CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SORT6"  
 <1. 1493  
 /codon\_start=3  
 /product="Similar to N-acylsphingosine amidohydrolase 2"  
 /protein\_id="AAH22604.1"  
 /db\_xref="GI:18490447"  
 /translation="KLVPLNGEDJGLISMFALHPYSNNKSNHFFVSDNMGYATLFEQ  
 EKNKYLPGGQPFVAFASNSLGDVSPNLGHCYVNTGSCNDSTCCNGSPSCMA  
 SGPGDMFESTHIIIGRIIYOKAKELIYASASQVTPVLAHQVMWMTDVSVDLNAHT  
 VKTKPALGYSFAAGTIDVSGSLNTIGCTTEBDPFWDLRLDLKPSSEIVECQPK

BASE COUNT 467 a 355 c 370 g 379 t  
 ORIGIN  
 PILLHSELTPHPMPDIDVDIVIVTGSILAIAPGELTTMSGRFREARKKEFALY  
 GMDMTVVASNSNYTHITTYEEEOARYSRASYIYGPHTLSAVIOLEFRLAKAIA  
 TPTVAANVSGSPPKPKNLIALSLINADRIAGHFGVLOPATREYGVVAVIF  
 VGANRNSAENOTHOTELTYEYEDSVADMOYNDASBETRFYHKGILGISMNTIY  
 WHIPDAIPGIRIRFGRHNRROELKPAVILAFBISSEFVVTT"  
 Query Match 72.9%; Score 1493; DB 10; Length 1571;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 557 TGAATGCTGATTTGAATGAGAAAGACTTGGCTTTCATGCTGTTGCATCCACC 616  
 Db 1 TGAATGCTGATTTGAATGAGAAAGACTTGGCTTTCATGCTGTTGCATCCACC 60  
 QY 617 CCGTAGCATGAACAATATAGCAACACCTTTGTAAATAGTACATATAGGCTATGCGGCTT 676  
 Db 61 CCGTAGCATGAACAATATAGCAACACCTTTGTAAATAGTACATATAGGCTATGCGGCTT 120  
 QY 677 ACCCTTTTGAAGAAAAGAAAGCAAGCTATCTGCTGAGAGAGACCGCTTTGTAGCAG 736  
 Db 121 ACCCTTTTGAAGAAAAGAAAGCAAGCTATCTGCTGAGAGAGACCGCTTTGTAGCAG 180  
 QY 737 GCTTTCCTTCATCAAAATCTCGAGAGAGCTGACCCCAACATCTTGGCCCGCATTTGTCTA 796  
 Db 181 GCTTTCCTTCATCAAAATCTCGAGAGAGCTGACCCCAACATCTTGGCCCGCATTTGTCTA 240  
 QY 797 ACACAGGGAGTCTTGTGACAAACAGACAGAGACACCTGTCCACAGGTGGGCTAGCATGT 856  
 Db 241 ACACAGGGAGTCTTGTGACAAACAGACAGAGACACCTGTCCACAGGTGGGCTAGCATGT 300  
 QY 857 GCATGGCCAGCGAGCTGTGACAAAGACATTTTGAAGAGACACACANTATAGAGGATCA 916  
 Db 301 GCATGGCCAGCGAGCTGTGACAAAGACATTTTGAAGAGACACACANTATAGAGGATCA 360  
 QY 917 TCTATCAGAAGGCCAAGAGAGCTGTATGCTTGCCTCCAGAGAGGTGACCGGCCAGTGC 976  
 Db 361 TCTATCAGAAGGCCAAGAGAGCTGTATGCTTGCCTCCAGAGAGGTGACCGGCCAGTGC 420  
 QY 977 TTGCAGCTCACCAGTGGGTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCCACACA 1036  
 Db 421 TTGCAGCTCACCAGTGGGTGAACATGACAGATGTGAGCGTCCAGCTCAATGCCCACACA 480  
 QY 1037 CAGTGAAGACGTGTAAACCTGCGCTGGCTACAGTTTCCGACAGGCAATGTATGAGG 1096  
 Db 481 CAGTGAAGACGTGTAAACCTGCGCTGGCTACAGTTTCCGACAGGCAATGTATGAGG 540  
 QY 1097 TTTGGGCTCCTCAATATTACACAGGGAACCTAGGAGGAGATCCATTCTGGACACTCTTC 1156  
 Db 541 TTTGGGCTCCTCAATATTACACAGGGAACCTAGGAGGAGATCCATTCTGGACACTCTTC 600  
 QY 1157 GGGACACAGCTTTGGGAAAACCATCTGAAGAGATTGTAGAGTGCAGAAACCAACCA 1216  
 Db 601 GGGACACAGCTTTGGGAAAACCATCTGAAGAGATTGTAGAGTGCAGAAACCAACCA 660  
 QY 1217 TCCCTGTTACAGCTGAGAGAGCTGACATACCATCTTGGACACAGATATTGTGTAGG 1276  
 Db 661 TCCCTGTTACAGCTGAGAGAGCTGACATACCATCTTGGACACAGATATTGTGTAGG 720  
 QY 1277 TTCAGATGTTACCGTTGGGCTCTTGGCCATAGCTGCTTATCCCTGGGGAATTAACACCA 1336  
 Db 721 TTCAGATGTTACCGTTGGGCTCTTGGCCATAGCTGCTTATCCCTGGGGAATTAACACCA 780  
 QY 1337 TGTGGGACGAAGATTTGCTGAGGCAATTTAAAAAGAAATTTGCCTTATAGGATGAGG 1396  
 Db 781 TGTGGGACGAAGATTTGCTGAGGCAATTTAAAAAGAAATTTGCCTTATAGGATGAGG 840  
 QY 1397 AATATGACCTGTTATTCGAGAGCTTAAAGAAATTTATACATTAACATTAACATTAAG 1456  
 Db 841 AATATGACCTGTTATTCGAGAGCTTAAAGAAATTTATACATTAACATTAACATTAAG 900  
 QY 1457 AAGAAATACAGAGCTCAGCGTACGAGGCGAGCTTACATCTATGAGACACACACCTGT 1516

```

Db      901  |MAGATATCCAGGCTTCAGCGGTACGAGGACGACATCTACAACTTTGACCCACACACCTGT 960
Oy      1517 |CTGCATACATCAACACTCTTCAGAGACCTTGCTAAGGCAATTCGTACGACACAGTACCA 1576
Db      961  |CTGCATACATCAACACTCTTCAGAGACCTTGCTAAGGCAATTCGTACGACACAGTACCA 1020
Oy      1577 |ACATGAGCACTGTCGCCGAGCTTCATCTTCAAAAATCTAATAGCTTCCTACTATTCCTA 1636
Db      1021 |ACATGAGCACTGTCGCCGAGCTTCATCTTCAAAAATCTAATAGCTTCCTACTATTCCTA 1080
Oy      1637 |ATATTGGGATAGAGCAACCAATTTGGGAGATGCTTGCCAGCCAGCAAAAC 1696
Db      1081 |ATATTGGGATAGAGCAACCAATTTGGGAGATGCTTGCCAGCCAGCAAAAC 1140
Oy      1697 |CTGAATACAGAGTGGGAGAGAGTGGTGAAGTATATTTTGAAGCGCTTACCCAAAGATT 1756
Db      1141 |CTGAATACAGAGTGGGAGAGAGTGGTGAAGTATATTTTGAAGCGCTTACCCAAAGATT 1200
Oy      1757 |CAGCAGAGAACCGAGACCCATCAAACTTCTCAGTGGAGAAATACGAGAGCTCTGTAG 1816
Db      1201 |CAGCAGAGAACCGAGACCCATCAAACTTCTCAGTGGAGAAATACGAGAGCTCTGTAG 1260
Oy      1817 |CTGACTGGCAGATATATATATACGATGCTCTGCGGAGAGAGGTTTATTGGACAAAG 1876
Db      1261 |CTGACTGGCAGATATATATATACGATGCTCTGCGGAGAGAGGTTTATTGGACAAAG 1320
Oy      1877 |GAATACTGGGCTGAGCAATGCAACAATATATCTGATATTCAGATACTGCTTACCTG 1936
Db      1321 |GAATACTGGGCTGAGCAATGCAACAATATATCTGATATTCAGATACTGCTTACCTG 1380
Oy      1937 |GAATCTACAGATATATATTTTGGACACATGCGAAGCAAGACTTTCGAAACCCGCTG 1996
Db      1381 |GAATCTACAGATATATATTTTGGACACATGCGAAGCAAGACTTTCGAAACCCGCTG 1440
Oy      1997 |TCATCTACAGATTTGAAGGAATTTCTCTCTTTGAGTGTGACTACTAG 2049
Db      1441 |TCATCTACAGATTTGAAGGAATTTCTCTCTTTGAGTGTGACTACTAG 1493

RESULT 4
AC117816 165407 bp DNA linear HTG 16-AUG-2002
LOCUS Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE, 12
DEFINITION
unordered pieces.
AC117816
VERSION AC117816.2 GI:22267711
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 165407)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-573C13
unpublished
2 (bases 1 to 165407)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hago,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,J., Levine,R., Larocque,K., Lamazares,R.,
Lander,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,C.,
Michova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connell,P., O'Neill,D., O'Neill,J.,
Oliker,J., Peterson,K., Phunkhang,P., Plerre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seamus,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Stauss,N., Subramanian,A., Talmas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 165407)
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hago,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Menus,C., Mlenga,V., Norbu,C., Norman,C.H.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Plerre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seamus,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talmas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 16, 2002 this sequence version replaced gi:20128606.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 573C13
Center clone name: 573C13
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160616 bases at least Q40
Consensus quality: 162967 bases at least Q30
Consensus quality: 163827 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 164307; sum-of-ctrls
Quality coverage: 7.1 in Q20 bases; agarose-fp
Quality coverage: 7.2 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
705: contig of 705 bp in length
706 805: gap of 100 bp
806 1769: contig of 984 bp in length
1790 1889: gap of 100 bp
1890 3396: contig of 1507 bp in length
3397 3496: gap of 100 bp
3497 4739: contig of 1243 bp in length
4740 4839: gap of 100 bp
4840 11073: contig of 6234 bp in length
11074 11173: gap of 100 bp

```



```

* 11174 18734: contig of 7561 bp in length
* 18735 18834: gap of 100 bp
* 18835 29191: contig of 10357 bp in length
* 29192 29291: gap of 100 bp
* 29292 42552: contig of 13261 bp in length
* 42553 42652: gap of 100 bp
* 42653 60018: contig of 17366 bp in length
* 60019 60118: gap of 100 bp
* 60119 81529: contig of 21411 bp in length
* 81530 81629: gap of 100 bp
* 81630 120500: contig of 38871 bp in length
* 120501 120600: gap of 100 bp
* 120601 165407: contig of 44807 bp in length.

```

## FEATURES

Location/Qualifiers

```

1. 165407
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_RP24-573C13"
/clone_lib="RPC1-24 Male Mouse BAC"
1. 705

```

```

misc_feature /note="assembly_fragment" 806..1789
misc_feature /note="assembly_fragment" 1890..3396
misc_feature /note="assembly_fragment" 3497..4739
misc_feature /note="assembly_fragment" 4840..11073
misc_feature /note="assembly_fragment" 11174..18734
misc_feature /note="assembly_fragment" 18835..29191
misc_feature /note="assembly_fragment" 29292..42552
misc_feature /note="assembly_fragment" 42653..60018
misc_feature /note="assembly_fragment" 60119..81529
misc_feature /note="assembly_fragment" 81630..120500
misc_feature /note="assembly_fragment" 120601..165407
misc_feature /note="assembly_fragment" 165407..339090
BASE COUNT 46319 a 32936 c 33090 g 51958 t 1104 others
ORIGIN

```

```

Query Match 9.48; Score 192; DB 2; Length 165407;
Best Local Similarity 100.0%; Pred. No. 2.1e-103;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1858 AGTTTATTTGGCAAGAAAGATCTGGCTGAGCAATGCAACATTTACTGGCATATT 1917
DB 1453 AGGTTTATTTGGCAAGAAAGATCTGGCTGAGCAATGCAACATTTACTGGCATATT 1512
OY 1918 CCAGATACCTGCTACCTCGAATCTACAGAAATAGATTTTGGACACATCGAAGCAG 1977
DB 1513 CCAGATACCTGCTACCTCGAATCTACAGAAATAGATTTTGGACACATCGAAGCAG 1572
OY 1978 GAATCTTGAACCCCGCTGATACACTGATTTGAGAAATTTTCTCTTGGAGATT 2037
DB 1573 GAATCTTGAACCCCGCTGATACACTGATTTGAGAAATTTTCTCTTGGAGATT 1632
OY 2038 GTCACTACTTAG 2049
DB 1633 GTCACTACTTAG 1644

```

```

RESULT 5
AC117816 165407 bp DNA linear HTG 16-AUG-2002
LOCUS Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE, 12
DEFINITION
ACCESSION AC117816
unsorted pieces.

```

## VERSION

AC117816.2 GI:22267711  
HTG: HTGS\_PHASE1, HTGS\_DRAFT.  
SOURCE  
house mouse.  
Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 165407)

## TITLE

Birren,B., Nusbaum,C. and Lander,E.

## JOURNAL

Mus musculus, clone RP24-573C13  
Unpublished  
2 (bases 1 to 165407)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Baran,N., Bastien,J., Bloom,T., Boguslavsky,L., Bouckgalter,B., Brown,A., Camarata,J., Campolongo,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,D., Gardyna,S., Glinde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

Direct Submission  
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 165407)

## REFERENCE

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Baran,N., Bastien,J., Bloom,T., Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,D., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## JOURNAL

Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 16, 2002 this sequence version replaced g1:20128606.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

## AUTHORS

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: 573\_C13  
Center clone name: 125854  
Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960721  
 Consensus quality: 160616 bases at least Q40  
 Consensus quality: 162967 bases at least Q30  
 Consensus quality: 163827 bases at least Q20  
 Insert size: 168000; agarose-fp  
 Insert size: 164307; sum-of-coverage  
 Quality coverage: 7.1 in Q20 bases; agarose-fp  
 Quality coverage: 7.2 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 705: contig of 705 bp in length  
 706 805: gap of 100 bp  
 806 1789: contig of 984 bp in length  
 1790 1889: gap of 100 bp  
 1890 3396: contig of 1507 bp in length  
 3397 3496: gap of 100 bp  
 3497 4739: contig of 1243 bp in length  
 4740 4839: gap of 100 bp  
 4840 11073: contig of 6234 bp in length  
 11074 11173: gap of 100 bp  
 11174 18734: contig of 7561 bp in length  
 18735 18834: gap of 100 bp  
 18835 29191: contig of 10357 bp in length  
 29192 29291: gap of 100 bp  
 29292 42552: contig of 13261 bp in length  
 42553 42652: gap of 100 bp  
 42653 60018: contig of 17366 bp in length  
 60019 60118: gap of 100 bp  
 60119 81529: contig of 21411 bp in length  
 81530 81629: gap of 100 bp  
 81630 120500: contig of 38871 bp in length  
 120501 120600: gap of 100 bp  
 120601 165407: contig of 44807 bp in length.

## FEATURES

source

1. 165407  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_id="RP24-573C13"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 1. 705  
 /note="assembly-fragment"  
 806 1789  
 /note="assembly-fragment"  
 1890 3396  
 /note="assembly-fragment"  
 3497 4739  
 /note="assembly-fragment"  
 4840 11073  
 /note="assembly-fragment"  
 11174 18734  
 /note="assembly-fragment"  
 18835 29191  
 /note="assembly-fragment"  
 29292 42552  
 /note="assembly-fragment"  
 42653 60018  
 /note="assembly-fragment"  
 60119 81529  
 /note="assembly-fragment"  
 81630 120500  
 /note="assembly-fragment"  
 120601 165407  
 /note="assembly-fragment"  
 misc\_feature 46319 a 32936 c 33090 g 51958 t 1104 others  
 BASE COUNT  
 ORIGIN

Query Match 8.7%; Score 179; DB 2; Length 165407;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-95;  
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

215 AGGTCTGAAGAGACATGAGAGTAATATGCGTCTGTATGAGAGAGCAATGTTATCC 274  
 Db 17534 AGGTCTGAAGAGACATGAGAGTAATATGCGTCTGTATGAGAGAGCAATGTTATCC 17415  
 QY 275 TGAGTCCATTGACACACACTCTGGCCAGCAGAGGTTTTCATATATACATCTATATAC 334  
 Db 17474 TGAGTCCATTGACACACACTCTGGCCAGCAGAGGTTTTCATATATACATCTATATAC 17415  
 QY 335 TCGCCAGCAGGAGGATTACACACCGGACCTTTTCAGTACATCTCTGAGATCATGAG 333  
 Db 17414 TCGCCAGCAGGAGGATTACACACCGGACCTTTTCAGTACATCTCTGAGATCATGAG 17356

## RESULT 6

AC113485

LOCUS 184101 bp DNA linear HTG 06-JUN-2002  
 DEFINITION Mus musculus clone RP23-346D12, WORKING DRAFT SEQUENCE, 17 ordered pieces.

AC113485

AC113485.3 GI:21327530  
 HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS house mouse.

SOURCE

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 184101)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faru, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., McKernan, K., McElrath, N., Meneus, L., Milnova, T., Meneus, L., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Olvera, J., Peterson, K., Phunkhang, P., Pierre, N., Polara, Y., Raymond, C., Retta, R., Riedack, M., Riley, R., Risse, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testafaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

REFERENCE

AUTHORS

REFERENCE

AUTHORS

Direct Submission  
 Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 184101)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faru, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., McKernan, K., McElrath, N., Meneus, L., Milnova, T., Meneus, L., Murphy, T., Naylor, J., Nguyen, C.,

TITLE  
JOURNAL

## COMMENT

Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 6, 2002 this sequence version replaced gi:21313862.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

## Project Information

Center project name: L23776

Center clone name: 346\_D\_12

## Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178331 bases at least Q40

Consensus quality: 181207 bases at least Q30

Consensus quality: 182003 bases at least Q20

Insert size: 183000; agarose-fp

Insert size: 182501; sum-of-contigs

Quality coverage: 6.7 in Q20 bases; agarose-fp

Quality coverage: 6.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced

by the finished sequence as soon as it is available and

the accession number will be preserved.

1 1186: contig of 1186 bp in length  
1187 1286: gap of 100 bp  
1287 2509: contig of 1223 bp in length  
2510 2609: gap of 100 bp  
2610 4200: contig of 1591 bp in length  
4201 4300: gap of 100 bp  
4301 6783: contig of 2483 bp in length  
6784 6883: gap of 100 bp  
6884 7958: contig of 1075 bp in length  
7959 8058: gap of 100 bp  
8059 10796: contig of 2738 bp in length  
10797 10896: gap of 100 bp  
10897 13878: contig of 2982 bp in length  
13879 13978: gap of 100 bp  
13979 15988: contig of 2010 bp in length  
15989 16088: gap of 100 bp  
16089 20675: contig of 4587 bp in length  
20676 20775: gap of 100 bp  
20776 25506: contig of 4731 bp in length  
25507 25606: gap of 100 bp  
25607 30961: contig of 5355 bp in length  
30962 31061: gap of 100 bp  
31062 41639: contig of 10578 bp in length  
41640 41739: gap of 100 bp  
41740 77584: contig of 35845 bp in length  
77585 77684: gap of 100 bp  
77685 107166: contig of 29482 bp in length  
107167 107266: gap of 100 bp

FEATURES  
source

\* 107267 140582: contig of 3336 bp in length  
\* 140583 140682: gap of 100 bp  
\* 140683 174914: contig of 34232 bp in length  
\* 174915 175014: gap of 100 bp  
\* 175015 184101: contig of 9087 bp in length.  
Location/Qualifiers  
1. 184101

## misc\_feature

/db\_xref="taxon:10090"  
/clone\_id="RPC1-23 Female Mouse BAC"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

## misc\_feature

/note="assembly\_fragment"

BASE COUNT 51100 a 39204 c 38438 g 53758 t 1601 others

## ORIGIN

## Query Match

Best Local Similarity 100.0%; Pred. No. 1.5e-95; Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## Query

215 AGGTCCTGGAAGACTAGAGAGTAATATGCTCTCTGATCGAAGACAATGTTTCC 274

## Db

17405 AGGTCCTGGAAGACTAGAGAGTAATATGCTCTCTGATCGAAGACAATGTTTCC 17464

## Query

275 TGAAGTCATTCACACACACTCTGCGCAGAGGTTTTCATATATACACTATATAC 334

## Db

17465 TGAAGTCATTCACACACACTCTGCGCAGAGGTTTTCATATATACACTATATAC 17524

## Query

335 TCGCCAGCAGAGGATTCAGCAACCGACCTTTCAGTACATAGTCTGGATCATGAAG 393

## Db

17525 TCGCCAGCAGAGGATTCAGCAACCGACCTTTCAGTACATAGTCTGGATCATGAAG 17583

## RESULT 7

## AC113485/c

## LOCUS

## DEFINITION

## Mus musculus

## clone

## RP23-346D12,

## WORKING

## DRAFT

## SEQUENCE,

## 17 ordered

## 184101 bp

## DNA

## linear

## HTG

## 06-JUN-2002

pieces.  
AC113485 GI:21327530  
HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
house mouse.  
SOURCE  
ORGANISM  
Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 184101)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barne, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Chang, J., Chazaro, B.,  
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
Lander, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,  
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Struss, P., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travs, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 184101)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barne, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,  
Chazaro, B., Chopel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,  
Lander, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,  
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travs, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 6, 2002 this sequence version replaced gi:21313862.  
ALL repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
Smit, A.F.A. & Green, P. (1996-1997)

Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L23776  
Center clone name: 346\_D.12  
----- Summary Statistics -----  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 17831 bases at least Q40  
Consensus quality: 181207 bases at least Q30  
Consensus quality: 182003 bases at least Q20  
Insert size: 183000; agarose-1p  
Insert coverage: 182501; sum-of-contigs  
Quality coverage: 6.7 in Q20 bases; agarose-1p  
Quality coverage: 6.7 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 1186: contig of 1186 bp in length  
1 1187 1286: gap of 100 bp  
1 1287 2509: contig of 1223 bp in length  
2510 2609: gap of 100 bp  
2610 4200: contig of 1591 bp in length  
4201 4300: gap of 100 bp  
4301 6783: contig of 2483 bp in length  
6784 6883: gap of 100 bp  
6884 7958: contig of 1075 bp in length  
7959 8058: gap of 100 bp  
8059 10796: contig of 2728 bp in length  
10797 10896: gap of 100 bp  
10897 13878: contig of 2992 bp in length  
13879 13978: gap of 100 bp  
13979 15988: contig of 2010 bp in length  
15989 16088: gap of 100 bp  
16089 20675: contig of 4587 bp in length  
20676 20775: gap of 100 bp  
20776 25506: contig of 4731 bp in length  
25507 25606: gap of 100 bp  
25607 30961: contig of 5355 bp in length  
30962 31061: gap of 100 bp  
31062 41639: contig of 10578 bp in length  
41640 41739: gap of 100 bp  
41740 77584: contig of 35845 bp in length  
77585 77684: gap of 100 bp  
77685 107166: contig of 29482 bp in length  
107167 107266: gap of 100 bp  
107267 140582: contig of 33316 bp in length  
140583 140682: gap of 100 bp  
140683 174914: contig of 34232 bp in length  
174915 175014: gap of 100 bp  
175015 184101: contig of 9087 bp in length.

#### FEATURES

source  
1. 184101  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-346D12"  
/clone\_11b="RPci-23 Female Mouse BAC"  
1. 1186  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
1287. 2509  
/note="assembly\_fragment"  
2610. 4200  
/note="assembly\_fragment"  
4301. 6783  
/note="assembly\_fragment"

```
misc_feature      6884..7958
                  /note="assembly_fragment"
misc_feature      8059..10796
                  /note="assembly_fragment"
misc_feature      10897..13878
                  /note="assembly_fragment"
misc_feature      13979..15988
                  /note="assembly_fragment"
misc_feature      16089..20675
                  /note="assembly_fragment"
misc_feature      20776..22506
                  /note="assembly_fragment"
misc_feature      23607..30961
                  /note="assembly_fragment"
misc_feature      31062..41639
                  /note="assembly_fragment"
misc_feature      41740..77584
                  /note="assembly_fragment"
misc_feature      77685..107166
                  /note="assembly_fragment"
misc_feature      107267..140582
                  /note="assembly_fragment"
misc_feature      140683..174914
                  /note="assembly_fragment"
misc_feature      175015..184101
                  /note="assembly_fragment"
misc_feature      clone_end:T7
                  vector_side:right"
BASE COUNT      51100 a 39204 c 38438 g 53758 t 1601 others
ORIGIN
Query Match      3.5%; Score 72; DB 2; Length 184101;
Best Local Similarity 100.0%; Pred. No. 7.3e-31;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      147 AGATGGTCAATCGAATGCGATTGTGAGCGTGAACATATATGATTTCCCAACGACT 206
         |||||||
Db       30961 AGATGGTCAATCGAATGCGATTGTGAGCGTGAACATATATGATTTCCCAACGACT 30962
QY      207 GAGGTTGAGGCT 218
         |||||||
Db       30901 GAGGTTGAGGCT 30890

RESULT 8
E50433
LOCUS      E50433      2283 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Novel protein.
ACCESSION  E50433
VERSION    E50433.1 GI:18633518
KEYWORDS  JP 2001057890-A/1.
SOURCE    Rattus norvegicus.
ORGANISM  Rattus norvegicus.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1 (bases 1 to 2283)
AUTHORS  Omori,A. and Ito,M.
TITLE    Novel protein.
JOURNAL  Patent: JP 2001057890-A 1 06-MAR-2001;
         MITSUBISHI CHEM CORP
COMMENT  OS Rattus norvegicus (rat)
         PN JP 2001057890-A/1
         PD 06-MAR-2001
         PF 23-AUG-1999 JP 1999235218
         PR
FEATURES
         key          Location/Qualifiers
         FT          CDS
                   Location/Qualifiers
                   (1)..(2283).
```

```
source      1..2283
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
BASE COUNT  641 a 584 c 524 g 534 t
ORIGIN
Query Match      3.3%; Score 68; DB 6; Length 2283;
Best Local Similarity 100.0%; Pred. No. 1.8e-28;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1057 GCCCTGGGCTACAGTCTTCCGACAGCACAATGATGAGTTCGGCCCTCAATATTACA 1116
         |||||||
Db       1294 GCCCTGGGCTACAGTCTTCCGACAGCACAATGATGAGTTCGGCCCTCAATATTACA 1353
QY      1117 CAGGGAAC 1124
         |||||||
Db       1354 CAGGGAAC 1361

RESULT 9
AB057433
LOCUS      AB057433      2456 bp      mRNA      linear      ROD 12-JUL-2001
DEFINITION Rattus norvegicus PAPKCD mRNA for ceramidase, complete cds.
ACCESSION  AB057433
VERSION    AB057433.1 GI:14701597
KEYWORDS
SOURCE    Rattus norvegicus kidney cDNA to mRNA.
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1
AUTHORS  Mitsuake,S., Tanii,M., Okino,N., Mori,K., Ichinose,S., Omori,A.,
         Iida,H., Nakamura,T. and Ito,M.
TITLE    Purification, characterization, molecular cloning, and subcellular
         distribution of neutral ceramidase of rat kidney
JOURNAL  J. Biol. Chem. 276 (28), 26249-26259 (2001)
MEDLINE  21365561
REFERENCE  2 (bases 1 to 2456)
AUTHORS  Ito,M.
TITLE    Direct Submission
JOURNAL  Submitted (14-MAR-2001) Makoto Ito, Kyushu University, Bioresource
         and Bioenvironmental Sciences; hakozaaki 6-10-1, higaashi-ku, Fukuoka
         812-8581, Japan (E-mail:makotoieagr.kyushu-u.ac.jp,
         Tel:81-92-642-2900, Fax:81-92-642-2900)
FEATURES
         source
         gene
         CDS
           /gene="PAPKCD"
           /gene="PAPKCD"
           /gene="PAPKCD"
           /codon_start=1
           /product="ceramidase"
           /protein_id="PAB62033.1"
           /db_xref="GI:14701598"
           /translation="MAKRTFSLEAFILFLVMTAIVALLLLFVTSGETIENHKDS
           GNMWSTTGGPTTQSPPTQPTTPDLPSPSNGFYIGVRACDCTQVSDINM
           GYKNGCONAGOLILRLFSRAFLIADPGSRNMAFVSVELCMIISRLLEVLKRSKY
           GSILRDNVLTLSATHTSGPAGFQVYLLIASGSPNRTFOYVSGIVASIDIAHN
           LKRGKVLINKGANVNOINRSPSYLONPSEARYSDDDKEMVYKIKVDINGEDG
           LISFPAVHPVSMNSNHLVSDNNGTAATLPEDEKNRGYLPQGPFPAGRASNLGCV
           SPNLGPHCVNTGESCDNRSTCEPSGSPSCMASPGQDMFESTHIIIGRYIVOKREL
           HASASQEVTPGVLTIAHQVMNMIDVSVOLNATHVKTCKAALGYSFAGTIDGVSKML
           TQGTTEGNLFMDTLRDOLGKPSSEIIECQKPKPILITGELTLPHPMOPDIYDIQIV
           TLGSLATAIIPGETTWSGRRLRAVKKERALKGMKMTVIVGLSNVYHYITTYE
           KOARVYAASTIYCPHTLSAYIOLFRALAKAINTPYANNSGPEPEFKNLIGSLIP
           NIADRAPIGRQFGDVLDPAPKETRVGVEVVEYGANPKSAENQHTOTLTVEKIED
           SVANWQIMHNDASWETREYVHKVGLSNATIHMHIPDTALPGVYIRIRYGNHKKDEL
           LKPAVILAFEGISSPEIIVTT"
BASE COUNT  704 a 612 c 562 g 578 t
```

ORIGIN

Query Match 3.3%; Score 68; DB 10; Length 2456;  
Best Local Similarity 100.0%; Pred. No. 1.8e-28;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1057 GCCCTGGGCTACAGTTTGGCCGAGCACAATTGATGAGTTTCGGGCTCAATATTACA 1116  
|||||  
DB 1363 GCCCTGGGCTACAGTTTGGCCGAGCACAATTGATGAGTTTCGGGCTCAATATTACA 1422  
|||||

QY 1117 CAGGGAC 1124  
|||||

DB 1423 CAGGGAC 1430

RESULT 10  
E50435 539 bp DNA 1linear PAT 31-JAN-2002  
LOCUS E50435 Novel protein.  
DEFINITION E50435.  
ACCESSION E50435.1 GI:18633520  
VERSION JP 2001057890-A/3.  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE 1 (bases 1 to 539)  
AUTHORS Omori, A. and Ito, M.  
TITLE Novel protein  
JOURNAL Patent: JP 2001057890-A 3 06-MAR-2001;  
MITSUBISHI CHEM CORP  
COMMENT OS Artificial Sequence  
PN JP 2001057890-A/3  
PD 06-MAR-2001  
PF 23-AUG-1999 JP 1999235218  
PR  
PI AKIRA OMORI, MAKOTO ITO  
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/80, PC  
C12N15/00, C12N5/00  
CC  
FT key 1. Location/Qualifiers  
FT source 1. 539  
FT 1. Location/Qualifiers  
FEATURES  
source 1. 539  
Location/Qualifiers  
/organism="synthetic construct"  
/db\_xref="taxon:32630"

BASE COUNT 181 a 110 c 113 g 135 t

ORIGIN

Query Match 2.6%; Score 53; DB 6; Length 539;  
Best Local Similarity 100.0%; Pred. No. 2.1e-19;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1765 AACGAGACCATCAACCTCTCTCACTGTGAGAAATAGAGAGACTCTGAGC 1817  
|||||  
DB 1354 AACGAGACCATCAACCTCTCTCACTGTGAGAAATAGAGAGACTCTGAGC 206  
|||||

RESULT 11  
HUMD4C12M5 235 bp mRNA 1linear PRI 04-FEB-1999  
LOCUS HUMD4C12M5  
DEFINITION Human HepG2 partial cDNA, clone hmd4c12m5.  
ACCESSION D17062  
VERSION D17062.1 GI:598818  
KEYWORDS gene signature.  
SOURCE Homo sapiens Male cell\_line:HepG2 cDNA to mRNA, clone\_11b:Kiseru.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 235)  
AUTHORS Matoba, R., Okubo, K., Horii, N., Fukushima, A. and Matsubara, K.  
TITLE The addition of 5'-coding information to a 3'-directed cDNA library  
Improves analysis of gene expression

JOURNAL Gene 146 (2), 199-207 (1994)  
MEDLINE 94357437  
REFERENCE 2 (bases 1 to 235)  
AUTHORS Matoba, R.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUL-1993) Ryo Matoba, Osaka University, Institute for  
Molecular and Cellular Bio; 1-3, Yamada-oka, Suita, Osaka 565,  
Japan (E-mail:matoba@inherit.imb.osaka-u.ac.jp,  
Tel:81-6-877-5111 (ex. 3314), Fax:81-6-877-1922)  
Submitted (21-Jul-1993) to DDBJ by:  
Ryo Matoba  
COMMENT Molecular Microbiology and Genetics Lab.  
Research Institute of Innovative Technology for the Earth 9-2  
Kizugawadal Kizu-cyo,  
Soraku-gun, Kyoto  
Japan, 619-02  
Phone: 07747-5-2308  
Fax: 07747-5-2321.

FEATURES  
source 1. 235  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="Male"  
/cell\_line="HepG2"  
/clone\_11b="Kiseru"

BASE COUNT 64 a 46 c 57 g 68 t

ORIGIN

Query Match 2.3%; Score 47; DB 9; Length 235;  
Best Local Similarity 100.0%; Pred. No. 8.6e-16;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1933 CCTGAACTACAGATAGATATTTGGACACAAATCGAGACAGA 1979  
|||||  
DB 102 CCTGAACTACAGATAGATATTTGGACACAAATCGAGACAGA 148  
|||||

RESULT 12  
AF250847 2289 bp mRNA 1linear PRI 17-JUL-2000  
LOCUS AF250847  
DEFINITION Homo sapiens mitochondrial ceramidase mRNA, complete cds; nuclear  
gene for mitochondrial product.  
ACCESSION AF250847  
VERSION AF250847.1 GI:9246992  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2289)  
AUTHORS El Bawab, S., Roddy, P., Qian, T., Bielawska, A., Lemasters, J.J. and  
Hannun, Y.A.  
TITLE Molecular cloning and characterization of a human mitochondrial  
ceramidase  
JOURNAL J. Biol. Chem. 275 (28), 21508-21513 (2000)  
MEDLINE 20347271  
PUBMED 10781606  
REFERENCE 2 (bases 1 to 2289)  
AUTHORS El Bawab, S. and Hannun, Y.A.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2000) Biochemistry, Medical University of South  
Carolina, 114 Doughty Street, P.O. Box 250760, Charleston, SC  
29425, USA

FEATURES  
source 1. 2289  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1. 2286  
/function="cleaves lipid ceramide into free fatty acid and  
sphingosine"  
/note="active in neutral to alkaline range"  
/codon\_start=1  
/product="mitochondrial ceramidase"

```

/protein_id="AAF86240.1"
/db_xref="GI:9246993"
/translation="MSAIVALLSLFTTGTIEHKKDLGGHFFSTTOSPPATGOSTA
AORSTATOHSTATOSSTATOTSPVLPPESELPFONESGYHIGVRADCTGVADINLM
GYKSGONAGSILTRLYSRATIMAEPPDSNRTVEVDIGAVSOLRLIEVNRLOSKI
GSLRRDNVLLSTGTHSGPAGYFOYTVVFLASGFSNOTFOHMTGLIKLSIDIAHTN
MKPKRIFLNKGNVGVQIINRSPSYLONPOSEARYPENITDKEMIVLKMVDLNGDGL
LISWFAIHPVSMNNSNHLVNSDNGVYASYLEOEKRNKGYLEGQEPVAFASNSGDV
SPNLLGPCINTGSCDNANSTCPDIGGSMVCIARGQDMDSQIIGRAMVORAKEL
YASASOETGPIASAHQWMDTVLNSNTHASKTCKPALGYSPAACTIDGVCGLNF
TGGKTEGDPEDDTIRDOILGKPESEIEKCHKPKILLTGHSLKRPHPHPIDVQII
TLGSLAIIAIPGEFTTMSGRRLREAVOEFASHOMQNTVVISGLCNVYTHITTYEE
YQOARYEASTIYGPALISAYIOLFRNLAKAIADTVANLSIGSEPPFKDILYPLIP
SIVDRAPRGRTGVDVLOPAKPEYVGAETVIFGANKNKSNVONOTHOTFLVEKYEA
TSTSMOIVCNDASWETRRYMHKGLLSNATVEWHIPIPTAQIYIRIRYFGHNKKODI
LKPAVILISFEETSAPFEVYTI"

BASE COUNT      664 a      553 c      515 g      557 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 8,9e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCGGAATCTACAGATAAGATATTGTCGACCAATCGAAGCAGCA 1979
Db 2170 CCTGGAATCTACAGATAAGATATTGTCGACCAATCGAAGCAGCA 2216

RESULT 13
AX207121 2507 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 1 from Patent WO0155410.
DEFINITION AX207121
ACCESSION AX207121
VERSION AX207121.1 GI:15394925
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2507)
AUTHORS Hannun,Y.A. and el Bawab,S.
TITLE Ceramide compositions and methods based thereon
JOURNAL Patent: WO 0155410-A 1 02-AUG-2001;
MUSC Foundation For Research Development (US)
FEATURES
source
1. 2507
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
39..2324
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60194.1"
/db_xref="GI:15394926"
/translation="MSAIVALLSLFTTGTIEHKKDLGGHFFSTTOSPPATGOSTA
AORSTATOHSTATOSSTATOTSPVLPPESELPFONESGYHIGVRADCTGVADINLM
GYKSGONAGSILTRLYSRATIMAEPPDSNRTVEVDIGAVSOLRLIEVNRLOSKI
GSLRRDNVLLSTGTHSGPAGYFOYTVVFLASGFSNOTFOHMTGLIKLSIDIAHTN
MKPKRIFLNKGNVGVQIINRSPSYLONPOSEARYPENITDKEMIVLKMVDLNGDGL
LISWFAIHPVSMNNSNHLVNSDNGVYASYLEOEKRNKGYLEGQEPVAFASNSGDV
SPNLLGPCINTGSCDNANSTCPDIGGSMVCIARGQDMDSQIIGRAMVORAKEL
YASASOETGPIASAHQWMDTVLNSNTHASKTCKPALGYSPAACTIDGVCGLNF
TGGKTEGDPEDDTIRDOILGKPESEIEKCHKPKILLTGHSLKRPHPHPIDVQII
TLGSLAIIAIPGEFTTMSGRRLREAVOEFASHOMQNTVVISGLCNVYTHITTYEE
YQOARYEASTIYGPALISAYIOLFRNLAKAIADTVANLSIGSEPPFKDILYPLIP
SIVDRAPRGRTGVDVLOPAKPEYVGAETVIFGANKNKSNVONOTHOTFLVEKYEA
TSTSMOIVCNDASWETRRYMHKGLLSNATVEWHIPIPTAQIYIRIRYFGHNKKODI
LKPAVILISFEETSAPFEVYTI"

BASE COUNT      720 a      582 c      567 g      638 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 8,9e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1933 CCGGAATCTACAGATAAGATATTGTCGACCAATCGAAGCAGCA 1979
Db 2208 CCTGGAATCTACAGATAAGATATTGTCGACCAATCGAAGCAGCA 2254

RESULT 14
AL450382/c
LOCUS Human DNA sequence from clone Rpl1-56A21 on chromosome 10, complete
sequence.
DEFINITION AL450382 162228 bp DNA linear PRI 21-JUN-2002
ACCESSION AL450382
VERSION AL450382.7 GI:21540029
KEYWORDS htc.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 162228)
AUTHORS Peck,A.
TITLE Direct Submission
JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Jun 23, 2002 this sequence version replaced gi:14148851.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rpl1-56A21 is from the library RPl1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

FEATURES
source
1. 162228
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rpl1-56A21"
/clone_lib="RPl1-11.1"
/codon_start=1
BASE COUNT      51458 a      32622 c      31866 g      46282 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 9,5e-16;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 62706 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACGAGA 62660

RESULT 15  
AL513306 176865 bp DNA linear HTG: 10-JUL-2001  
LOCUS  
DEFINITION Homo sapiens chromosome 10 clone RP11-532F4, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AL513306  
VERSION AL513306.10 GI:14530311  
KEYWORDS HTG: HTGS\_PHASE2; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 176865)  
AUTHORS Howden, P.  
TITLE Direct Submision  
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Jun 23, 2001 this sequence version replaced gi:13274890.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: bA532F4  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 176570 bases at least Q40  
Consensus quality: 176555 bases at least Q30  
Consensus quality: 176594 bases at least Q20  
Insert size: 176865; sum-of-ctrls  
Insert size: 176572; 6.5% error; agarose-fp  
Quality coverage: 9.23x in Q20 bases; sum-of-ctrls Quality  
coverage: 9.25x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 176865  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-532F4"  
/clone\_lib="RP11-11.2"  
1. 176865  
/note="assembly fragment: 04692"  
BASE COUNT 53967 a 31937 c 33758 g 57203 t  
ORIGIN

Query Match 2.3%; Score 47; DB 2; Length 176865;  
Best Local Similarity 100.0%; Pred. No. 9.5e-16;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACGAGA 1979  
DB 32904 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACGAGA 32950

RESULT 16  
AL589794 178899 bp DNA linear PRI 18-SEP-2001  
LOCUS  
DEFINITION Human DNA sequence from clone RP11-564C4 on chromosome 10, complete  
sequence.  
ACCESSION AL589794  
VERSION AL589794.7 GI:14330084

KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 178899)  
AUTHORS Peck, A.  
TITLE Direct Submision  
JOURNAL Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Jun 8, 2001 this sequence version replaced gi:13992363.  
COMMENT During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
<http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
RP11-564C4 is from the library RP11-11.2 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6  
This sequence is the entire insert of clone RP11-564C4 The true  
left end of clone RP11-532F4 is at 81721 in this sequence.  
Location/Qualifiers  
1. 178899  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-564C4"  
/clone\_lib="RP11-11.2"  
BASE COUNT 51555 a 36501 c 36785 g 54058 t  
ORIGIN

Query Match 2.3%; Score 47; DB 9; Length 178899;  
Best Local Similarity 100.0%; Pred. No. 9.5e-16;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1933 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACGAGA 1979  
DB 114624 CCTGGAATCTACAGATAAGATATTTGGACACATCGAGACGAGA 114670

RESULT 17  
AC055726.2  
WPCOMMENT  
Sequence split into 5 fragments  
Fragment Name Begin End  
AC055726.0 1 110000  
AC055726.1 100001 210000  
AC055726.2 200001 310000  
AC055726.3 300001 410000  
AC055726.4 400001 456720  
Continuation (3 of 5) of AC055726 from base 200001 (AC055726 Homo sapiens chromosome  
10, complete sequence).  
Query Match 1.9%; Score 39; DB 2; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 6.5e-11;





```

* 30224 32496: contig of 2273 bp in length
* 32497 32596: gap of unknown length
* 32597 34038: contig of 1442 bp in length
* 34039 34138: gap of unknown length
* 34139 36429: contig of 2291 bp in length
* 36430 36529: gap of unknown length
* 36530 38289: contig of 1770 bp in length
* 38300 38399: gap of unknown length
* 38400 40786: contig of 2387 bp in length
* 40787 40886: gap of unknown length
* 40887 42807: contig of 1921 bp in length
* 42808 42907: gap of unknown length
* 42908 45554: gap of 2647 bp in length
* 45555 45654: gap of unknown length
* 45655 47282: contig of 1628 bp in length
* 47283 47382: gap of unknown length
* 47383 49177: contig of 1795 bp in length
* 49178 49277: gap of unknown length
* 49278 51282: contig of 2015 bp in length
* 51293 51392: gap of unknown length
* 52978 52978: contig of 1586 bp in length
* 52979 53078: gap of unknown length
* 53079 55113: contig of 2035 bp in length
* 55114 55213: gap of unknown length
* 55214 56557: contig of 1344 bp in length
* 56558 56657: gap of unknown length
* 56658 59123: contig of 2466 bp in length
* 59124 59223: gap of unknown length
* 59224 61914: contig of 2691 bp in length
* 61915 62014: gap of unknown length
* 62015 65172: contig of 3158 bp in length
* 65173 65272: gap of unknown length
* 65273 67490: contig of 2218 bp in length
* 67491 67590: gap of unknown length
* 67591 71140: contig of 3550 bp in length
* 71141 71240: gap of unknown length
* 71241 73628: contig of 2388 bp in length
* 73629 73728: gap of unknown length
* 73729 76582: contig of 2854 bp in length
* 76583 76682: gap of unknown length
* 76683 79308: contig of 2626 bp in length
* 79309 79408: gap of unknown length
* 79409 84267: gap of 4858 bp in length
* 84267 84366: gap of unknown length
* 84367 89469: contig of 5103 bp in length
* 89470 89569: gap of unknown length
* 89570 94790: contig of 5221 bp in length
* 94791 94890: gap of unknown length
* 94891 99546: contig of 4656 bp in length
* 99547 99646: gap of unknown length
* 99647 105838: contig of 6192 bp in length.

```

FEATURES  
source 1..105838  
Location/Qualifiers

BASE COUNT 28706 a 19526 c 20397 g 28661 t 8548 others  
ORIGIN

Query Match 1.8%; Score 36; DB 2; Length 105838;  
Best Local Similarity 100.0%; Pred. No. 4.2e-09;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1633 TTCACCTATTCCTAATATTCGCGATAGACACCAAT 1658  
Db 69354 TTCACCTATTCCTAATATTCGCGATAGACACCAAT 69389

RESULT 19 AC105714 105838 bp DNA linear HTG 13-JUL-2002  
AC105714/C Rattus norvegicus clone CH230-142G16, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 48 unordered pieces.

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AC105714  
AC105714.2 GI:21736609  
HTG: HTGS\_PHASE1.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE AUTHORS

1 (bases 1 to 105838)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alshrocks,S.L., Amaraltinge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbier,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,  
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burkett,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guetara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Hornl,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudah,S.,  
Karissom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,Z., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louisedge,H.,  
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapue,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokoko,S., Ogun,M., Okunolu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojiboken,I., Rolfe,M., Ruiz,S., Savery,G.,  
Schier,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,  
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

## TITLE

JOURNAL  
Unpublished  
2 (bases 1 to 105838)

## REFERENCE AUTHORS

JOURNAL  
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 105838)

## REFERENCE AUTHORS

JOURNAL  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:18092937.

## COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project information  
Center project name: GNR  
Center clone name: CH230-142G16  
Summary Statistics

```

Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 57464 bases at least Q40
Consensus quality: 61201 bases at least Q30
Consensus quality: 64573 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1227 1226: contig of 1226 bp in length
1327 1326: gap of unknown length
1327 12590: contig of 1264 bp in length
2591 2690: gap of unknown length
2691 3839: contig of 1149 bp in length
3840 3939: gap of unknown length
3940 5328: contig of 1389 bp in length
5329 5429: gap of unknown length
5429 6597: gap of unknown length
6597 7891: contig of 1195 bp in length
7892 7991: gap of unknown length
7992 9246: contig of 1255 bp in length
9247 9347 10448: contig of 1102 bp in length
10449 10549: gap of unknown length
10549 11837: contig of 1289 bp in length
11838 11937: gap of unknown length
11938 13599: contig of 1662 bp in length
13600 13699: gap of unknown length
13700 15116: contig of 1417 bp in length
15117 15216: gap of unknown length
15217 16273: contig of 1057 bp in length
16274 16374: gap of unknown length
16374 17450: contig of 1077 bp in length
17451 17550: gap of unknown length
17551 18836: contig of 1286 bp in length
18837 18936: gap of unknown length
18937 20118: contig of 1182 bp in length
20119 20218: gap of unknown length
20219 21463: contig of 1245 bp in length
21464 21563: gap of unknown length
21564 22745: contig of 1182 bp in length
22746 22845: gap of unknown length
22846 23877: contig of 1032 bp in length
23878 23977: gap of unknown length
23978 25337: contig of 1360 bp in length
25338 25437: gap of unknown length
25438 26655: contig of 1218 bp in length
26656 26755: gap of unknown length
26756 28546: contig of 1791 bp in length
28547 28646: gap of unknown length
28647 30123: contig of 1477 bp in length
30124 30223: gap of unknown length
30224 32496: contig of 2273 bp in length
32497 32596: gap of unknown length
32597 34038: contig of 1442 bp in length
34039 34138: gap of unknown length
34139 36429: contig of 2291 bp in length
36430 36529: gap of unknown length
36530 38399: contig of 1770 bp in length
38399 40786: gap of unknown length
40787 40886: contig of 2387 bp in length
40887 42807: contig of 1921 bp in length
42808 42907: gap of unknown length

```

```

* 42908 45554: contig of 2647 bp in length
* 45555 45654: gap of unknown length
* 45655 47282: contig of 1628 bp in length
* 47283 47382: gap of unknown length
* 47383 49177: contig of 1795 bp in length
* 49178 49277: gap of unknown length
* 49278 51292: contig of 2015 bp in length
* 51293 51392: gap of unknown length
* 51393 52978: contig of 1586 bp in length
* 52979 53078: gap of unknown length
* 53079 53113: contig of 2035 bp in length
* 53114 55213: gap of unknown length
* 55214 56557: contig of 1344 bp in length
* 56558 56658: gap of unknown length
* 56659 59123: contig of 2466 bp in length
* 59124 59223: gap of unknown length
* 59224 61914: contig of 2691 bp in length
* 61915 62014: gap of unknown length
* 62015 65172: contig of 3158 bp in length
* 65173 65272: gap of unknown length
* 65273 67490: contig of 2218 bp in length
* 67491 67591: gap of unknown length
* 67591 71140: contig of 3350 bp in length
* 71141 71240: gap of unknown length
* 71241 73628: contig of 2388 bp in length
* 73629 73728: gap of unknown length
* 73729 76582: contig of 2854 bp in length
* 76583 76682: gap of unknown length
* 76683 79308: contig of 2626 bp in length
* 79309 84266: gap of unknown length
* 84267 84366: gap of unknown length
* 84367 89469: contig of 5103 bp in length
* 89470 89569: gap of unknown length
* 89570 94790: contig of 5221 bp in length
* 94791 94890: gap of unknown length
* 94891 99546: contig of 4656 bp in length
* 99547 99646: gap of unknown length
* 99647 105838: contig of 6192 bp in length.
Location/Qualifiers
source
1..105838
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-142G16"
BASE COUNT 28706 a 19526 c 20397 g 28661 t 8548 others
ORIGIN
Query Match 1.6%; Score 32; DB 2; Length 105838;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1732 TTGTAGGCGCTAACCAAGATTTCAGCAGA 1763
|||||
Db 52897 TTTGTAGGCGCTAACCAAGATTTCAGCAGA 52866
RESULT 20
AL583858 187130 bp DNA linear HTG 10-JUL-2001
AL583858/C
LOCUS
DEFINITION
Hom sapiens chromosome 10 clone RP11-78J11, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
AL583858.9 GI:14575341
VERSION
KEYWORDS
HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 187130)
AUTHORS
Brown,J.
DIRECT SUBMISSION
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

```

COMMENT requests: clonerequest@sanger.ac.uk  
On Jun 28, 2001 this sequence version replaced gi:14330076.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: bA78J11  
----- Summary Statistics  
Sequencing program: XGAP4; version 4.5  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 187080 bases at least Q40  
Consensus quality: 187130 bases at least Q30  
Consensus quality: 187130 bases at least Q20  
Insert size: 187130; sum-of-contigs  
Insert size: 175296; 20.9% error; agarose-  
Quality coverage: 7.08x in Q20 bases; sum-of-contigs Quality  
coverage: 7.69x in Q20 bases; agarose-  
-----  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 187130  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone\_1lb="RP11-78J11"  
/clone\_1lb="RP11-11.1"  
1. 187130  
/note="assembly-fragment:03017"  
misc-feature 62248 a 37239 c 35291 g 52352 t  
BASE COUNT  
ORIGIN

Query Match 1.5%; Score 30; DB 2; Length 187130;  
Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1682 TGCAGCCAGCAAAACCTGAATACAGTGG 1711  
4136 TGCAGCCAGCAAAACCTGAATACAGTGG 4107

RESULT 21  
AX207126 38 bp DNA linear PART 30-AUG-2001  
LOCUS Sequence 6 from Patent WO0155410.  
DEFINITION AX207126  
ACCESSION AX207126  
VERSION AX207126.1 GI:15394932  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE 1 (bases 1 to 38)  
AUTHORS Hannun, Y.A. and el Bawab, S.  
TITLE Ceramide compositions and methods based thereon  
JOURNAL Patent: WO 0155410-A 6 02-AUG-2001;  
MUSC Foundation for Research Development (US)  
Location/Qualifiers  
1. 38  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer"  
BASE COUNT 11 a 10 c 10 g 7 t  
ORIGIN

Query Match 1.2%; Score 25; DB 6; Length 38;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1682 TGCAGCCAGCAAAACCTGAATACAG 1706  
Db 14 TGCAGCCAGCAAAACCTGAATACAG 38

RESULT 22  
OSJN00174 164805 bp DNA linear PLN 17-JUL-2002  
LOCUS Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0058K23,  
DEFINITION complete sequence.  
ACCESSION AL662970  
VERSION AL662970.2 GI:21741993  
KEYWORDS HTG.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriarthroideae; Oryzaceae; Oryza.  
REFERENCE 1  
AUTHORS Han, B., Feng, Q., Huang, Y.C., Li, Y., Zhu, J.J., Zhao, Q., Hu, X.,  
Liu, Y.L., Mu, J., Yu, Z., Chen, L., Fan, D.L., Meng, Q.D., Zhang, L.,  
Lu, Y.Q., Yu, S.L., Liu, X.H., Lu, T.T., Zhang, Y.J., Lu, Y., Li, C.,  
Li, T., Zhang, Y., Hu, H., Jia, P.X., Qian, Y.M., Ying, R., Zhou, B.,  
Chen, Z.H., Hao, P., Zhang, L., Wu, M., Zhang, R.Q., Guan, J.P., Fu, G.,  
Wang, S.T., Ren, S.X., Lv, G., Lin, W., Gu, W.Q., Zhu, G.F., Tu, Y.F.,  
Jia, J., Yin, H.F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X.Y.,  
Shao, C.Y., Sun, Y., Hu, Q.P., Zhang, X.L., Zhang, W., Wang, L.J.,  
Ding, C.W., Sheng, H.H., Gu, J.L., Chen, S.T., Ni, L., Zhu, F.H. and  
Hong, G.F.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,  
Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,  
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:  
bhan@ncgr.ac.cn  
REMARK Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC  
clone: OSJNBa0058K23.  
COMMENT On Jul 12, 2002 this sequence version replaced gi:17998482.  
Web site: http://www.ncgr.ac.cn  
----- Summary Statistics  
Assembly program: phrap  
-----  
NOTE: This is a complete sequence.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Fgenesh (<http://www.softberry.com/>),  
genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM  
(<http://genemark.biology.gatech.edu/genemark/>), tRNAscan-SE (Sean  
Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the  
complete sequence against NCBI non-redundant protein database (nr)  
(<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.  
Location/Qualifiers  
1. 164805  
/organism="Oryza sativa"  
/variety="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="4"  
/clone\_1lb="OSJNBa0058K23"  
/clone\_1lb="CUG1-OSJNBa"  
BASE COUNT 47435 a 34481 c 35046 g 47843 t  
ORIGIN

Query Match 1.1%; Score 23; DB 8; Length 164805;  
Best Local Similarity 100.0%; Pred. No. 0.31;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1676 ATGCTTCGACGCAAAACCT 1698  
Db 160127 ATGCTTCGACGCAAAACCT 160149

RESULT 23  
AC012131 175330 bp DNA linear HTG 14-SEP-2000  
LOCUS

```
DEFINITION Homo sapiens clone RP11-5E11, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
ACCESSION AC012131
VERSION AC012131.5 GI:10122074
KEYWORDS HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 175330)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
TITLE Homo sapiens, clone RP11-5E11
REFERENCE 2 (bases 1 to 175330)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
TITLE Homo sapiens, clone RP11-5E11
AUTHORS
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Domingo,M., Donegan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karates,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Neldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 14, 2000 this sequence version replaced gi:18576235.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2849
Center clone name: 5_E_11
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Sequencing vector: Plasmid; n/a; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172203 bases at least Q40
Consensus quality: 173820 bases at least Q30
Consensus quality: 174439 bases at least Q20
Insert size: 157000; agarose-IP
Insert size: 175030; sum-of-coverage
Quality coverage: 7.4 in Q20 bases; agarose-IP
Quality coverage: 6.6 in Q20 bases.
NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
777769: contig of 77769 bp in length
77770 77869: gap of 100 bp
77870 79091: contig of 1222 bp in length
79092 79191: gap of 100 bp
79192 139104: contig of 59913 bp in length
139105 139204: gap of 100 bp
139205 175330: contig of 36126 bp in length.
Location/Qualifiers
1..175330
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="RP11-5E11"
/clone_lib="RPCT-11 Human Male BAC"
1..77769
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
77870..79091
/note="assembly_fragment"
79192..139104
/note="assembly_fragment"
139205..175330
/note="assembly_fragment"
clone_end:T7
vector_side:right"

BASE COUNT 52740 a 33512 c 34651 g 54127 t 300 others
ORIGIN
Query Match 1.1%; Score 23; DB 2; Length 175330;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

940 TATGCTCTGCTCCAGAGGT 962
|||||
169446 TATGCTCTGCTCCAGAGGT 169468

RESULT 24
LOCUS F193462S02 168 bp DNA linear PRI 12-NOV-2000
DEFINITION Pan troglodytes dipeptidyl carboxy peptidase 1 (DCP1) gene, exon 2.
ACCESSION AF193463
VERSION AF193463.1 GI:11138388
KEYWORDS
SEGMENTS 2 of 26
SOURCE Pan troglodytes.
ORGANISM Pan troglodytes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (bases 1 to 168)
AUTHORS Dufour,C., Casane,D., Denton,D., Wickings,J., Corvol,P. and
Jeunemaitre,X.
TITLE Human-chimpanzee DNA sequence variation in the four major genes of
the renin-angiotensin system
JOURNAL Genomics 69 (1), 14-26 (2000)
MEDLINE 20469400
PUBMED 11013071
REFERENCE 2 (bases 1 to 168)
AUTHORS Dufour,C., Corvol,P. and Jeunemaitre,X.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1999) INSERM U36, College de France, 3 Rue d'Ulm,
Paris 75005, France
Location/Qualifiers
1..168
/organism="Pan troglodytes"
/db_xref="taxon:9598"
1..168
/gene="DCP1"
/number=2
/number=2

BASE COUNT 34 a 51 c 58 g 25 t
ORIGIN
Query Match 1.1%; Score 22; DB 9; Length 168;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

922 CAGAGGCCAGAGCTGTATG 943
|||||
46 CAGAGGCCAGAGCTGTATG 67

RESULT 25
AR137383
```

LOCUS AR137383 4020 bp DNA linear PAT 16-JUN-2001  
DEFINITION Sequence 130 from patent US 6197505.  
ACCESSION AR137383  
KEYWORDS AR137383.1 GI:14478892  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 4020)  
AUTHORS Norberg,L.,Torjorn., Andersson,M.,Kristina. and  
Lindstrom,P.,Harry,Rutger.  
TITLE Methods for assessing cardiovascular status and compositions for  
use thereof  
JOURNAL Patent: US 6197505-A 130 06-MAR-2001;  
FEATURES  
source Location/Qualifiers  
1..4020  
/organism="unknown"  
BASE COUNT 857 a 1261 c 1174 g 728 t  
ORIGIN  
Query Match 1.1%; Score 22; DB 6; Length 4020;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 922 CAGAGGCCAGAGAGCTGTATG 943  
Db 317 CAGAGGCCAGAGAGCTGTATG 338  
RESULT 26  
LOCUS HUMA1CEB 4020 bp mRNA linear PRI 30-OCT-1994  
DEFINITION Human angiotensin I-converting enzyme mRNA, complete cds.  
ACCESSION J04144  
VERSION J04144.1 GI:178285  
KEYWORDS angiotensin converting enzyme; dipeptidyl carboxypeptidase.  
SOURCE Human endothelial cell, cDNA to mRNA, clones lambda-HEC1922,  
lambda-HEC2111, and lambda-CHD32.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 4020)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE Soubrier,F., Albenc-Gelas,F., Hubert,C., Allegri,J., John,M.,  
Tregear,G. and Corvol,P.  
Two putative active centers in human angiotensin I-converting  
enzyme revealed by molecular cloning  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9386-9390 (1988)  
MEDLINE 89071703  
PUBMED 2849100  
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by  
F.Soubrier 04-JAN-1989.  
FEATURES  
source Location/Qualifiers  
1..4020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="17q23"  
1..4020  
/gene="DCP1"  
1..109  
/note="DCP1"  
/note="angiotensin I-converting enzyme signal peptide"  
23..3943  
/gene="DCP1"  
/note="angiotensin I-converting enzyme precursor (EC  
3.4.15.1)"  
3..4..15..1"  
/codon\_start=1  
/protein\_id="AA51684.1"  
/db\_xref="GI:178286"  
/db\_xref="GDB:600-119-840"  
/translation="MGAASGRGPGILLPLPLLLPPOPALPDGPGNFSADAE  
GAQLFAQSYNSAEQVLFQSVASNAHNTITAENARROEPAALLSOEAEAGAK  
ELYEIMNFTDPOLRITIGAVRTIGSANITLAKROOYNALLSNRSRYSTKVCILN  
KTAICWSLDPDLTNILASSRSYAMLLFAMEGHNAAGTLPKLFYEDFTALSNEAYKOD

GTTTGATWRSYNSPTFEEDLEHLVQOLEPLYLNLHFAVRALHRRGDRYINLRGP  
IPAHLLGDMMAQSWENTYDMVVPFDPKALDVTSMLOQGNATMFRFAEEFTSLE  
LSMPPEWEGSMLEKPADREVVCASAMDFNKRDEIRKCTQVTDQLSTVHEH  
GHIOYLYQKDPVSLRGANGFHEALIGDVALSVSPENHAKTGLDRVTCNDEH  
INVLKMALEKTAFLPEGYLNDQMRGVSRSRPNRMMWVIRTKOGICPPVTR  
NETHFDAGAKRFPVNVTPYIRFVSFVLOFQHEALCKAGTEGELHOCDDIRSKAG  
AKURVYLAGSSRPQEVLLKDWADLDAOPLVLFQVOTMLODQNOQNEVLGMP  
EYQHPPLPDNPESIDLVTDVAEASKEVEYDRTSQVWNEAYEAMNMYNNITTE  
SKILLQNNQIANTHLKYGTQARKFDVNOLOMTTIRIKKYQDLERAPALPAQLEBY  
NKLILDMETTYSVATYCHPNSCLOLEPDLTVNMTSRKREYEDLLMAGEMRDAGRAI  
LOEYRVYVELINQAAKLNGYVADGSMRSMEYTPSLIEDLERLEFOELDYLNLAHV  
RALHRRYGAOHIINEGPYPAHLINMMAQTSNTYDLYVPPSA.PSMDTEAMTKOG  
WPRRMRKADPFTSLGLVYPPRPWKSMLKRTDGRVYCHASAMDFYNGKPRRI  
KQCTVNLLEDLVVAHHEMGIQYEMQYKDLVPALREGANPGEHEAIGDVALSVTPK  
HLASLNLSSSEGSDEHDINFLMKALDKAIIPESYLDQMRWRFVDSITREYNQ  
EWSMLRKQGLCPVPPTGDFDGAKRIPSSVPIRYEVSFTIIOFHFALCOAA  
GHTGPLKCDIYOSKEAGORLATFAKLGSRPSPAMOLITOPNNSASMSYKPL  
LDMRTRENLEHGEKIGMPQYNTPNARSSEGPLPSGSEFLGLDLDAQARVQWML  
LFLGIALLVATIGLSORLEFSTRSLHRSLSHGSPQSGSEVELHRS"  
110..3940  
/gene="DCP1"  
/product="angiotensin I-converting enzyme"  
BASE COUNT 857 a 1261 c 1174 g 728 t  
ORIGIN  
Query Match 1.1%; Score 22; DB 9; Length 4020;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 922 CAGAGGCCAGAGAGCTGTATG 943  
Db 317 CAGAGGCCAGAGAGCTGTATG 338  
RESULT 27  
LOCUS AX429555 4022 bp DNA linear PAT 21-JUN-2002  
DEFINITION Sequence 18 from Patent WO0226820.  
ACCESSION AX429555  
VERSION AX429555.1 GI:21540823  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE Cohen,D., Dengler,U.J., Finelli,A.L., Freuler,F., Konsolaki,M.,  
Reinhardt,M.W. and Zusman,S.  
Transgenic drosophila melanogaster expressing beta amyloid  
JOURNAL NOVARTIS ERFIND VERMALT GMBH (AT)  
NOVARTIS  
FEATURES  
source Location/Qualifiers  
1..4022  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 856 a 1262 c 1174 g 728 t 2 others  
ORIGIN  
Query Match 1.1%; Score 22; DB 6; Length 4022;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 922 CAGAGGCCAGAGAGCTGTATG 943  
Db 319 CAGAGGCCAGAGAGCTGTATG 340  
RESULT 28  
LOCUS A00914 4024 bp DNA linear PAT 28-JAN-1993  
DEFINITION H.sapiens gene for angiotensin conversion enzyme (Ace).  
ACCESSION A00914  
VERSION A00914.1 GI:14652

KEYWORDS angiotensin-converting enzyme; peptidyl dipeptidase I.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS Soubrier, F.  
TITLE NUCLEIC ACID CODING FOR THE HUMAN ANGIOGENSINE CONVERSION ENZYME  
(ACE), AND ITS APPLICATIONS PARTICULARLY FOR THE IN VITRO DIAGNOSIS  
OF HIGH BLOOD PRESSURE  
JOURNAL Patent: WO 9003435-A 18 05-APR-1990;  
Institut National de la Sante et de la Recherche Med  
location/Qualifiers  
FEATURES  
source  
1. .4024  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
23. .3943  
/EC\_number="3.4.15.1"  
/codon\_start=1  
/product="peptidyl dipeptidase 1"  
/protein\_id="CA00097.1"  
/db\_xref="GI:14653"  
/translation="MGAASGRGCGLLPPLLLPDPALADPGLQNGNSADAE  
GAQLFAOSYNSAEOVLFOSYAASMAHDNTITAEAROEBAALLSEFAWQOKAK  
ELYEPIWQNTFDPLRLRIGAVRTLGASNLPLARQOYNALLSMRSITYSTAKCLPN  
KTATCWSLDPDLTNIILASSRSYAMLFAWEGMNAAGIPLKPLYEDFTALSNAYKOD  
GFHTGAVRSWYNSPTFEDDLHLVQOLEPLYNLHAFVRAALHRRYGDYINLRGP  
IPALIGDMAOSWENIYDMVVPDPKPNLDVYSTMLQQGNATHMRVAEFTSLE  
LSPMPPEWEGSMLEKPADGREYVCHASANDFNKRPFRIKQCTRYMDOLSTVHNEM  
GHIOYLYQYKDLVSLRGANPGFHEAIGDLAVSTPEHLKIGLDRVTNDESD  
INYLKMALEKIAFLPPGYLVDMQRMVFGSGRPSPSNEEDMWYLRKYOGICPPVR  
NETHFDAGAKFHVNPYIRYFVSFVLQFHFALCKEYEGEPLHOCODYRSTKAG  
AKRKVLQAGSSRPMOEVLKDMVGLDADPLKYOPATYKOPATYKOPATYKOPATYKOP  
EYOWHPRLPNYPGIDLVNDEASKEFVEYEDTSOVVWNEVREANNVNTNTTET  
SKILLQKMOIANTHTLKYGOARKEFDYNQONTTKRIKKVODLERAPALPAOLEEY  
NKILLDMETTSVATVCHPNQSCQLDLEPLTNWATSRKEDILMAWEGMDKGRRI  
LOEPKYVELLINQARLVYDAGDSWRSWYEPSPSLDLERLFOELQPLYNLHAYV  
RRALHRYGAQHINLEGPILPAHLIGNMAQTSNITDLYVFPSPASMDTTEAMLKOG  
WTPPRMKEADDEFTSGILLPVPEFPMKSMLEKTPGREVYCHASAMPDENGKDPRI  
KOCCTVNLDELIVAHHEMGIQYTMQYKDPVALREGANPGFHEAIGDLAVLSTSTRK  
HLHSLNLSSBGSDEHDINFLKMAULKIAFLPEFSYLVDMQRMVFDGSLTKKNYNO  
EWSWLRKYOGICPPVPTQGDDEPGAKFHPSSVPYIRYFVSFIQFHFALCKEYEGEPL  
GHTEPLKDKCIYOSKEAGORLATAMKLGFSRPMPEAMQLITGQPNMSASAMLSEFKPL  
LDMLRTNELHGEKLGMPQYNMTPNSAISEGPIIDSGRVSFLGDLDAQOARVQGWILL  
LFLGIALVFTLIGLSORLFSIRHRSILRHSHGPFQSGSEVELRHS"  
BASE COUNT 857 a 1263 c 1175 g 729 t  
ORIGIN  
Query Match 1.1%; Score 22; DB (c); Length 4024;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 922 CAGAGGCCAAGGAGCTGTATG 943  
Db 317 CAGAAGGCCAAGGAGCTGTATG 338  
|||||  
RESULT 29  
AF118569 24070 bp DNA linear PRI 08-AUG-2000  
LOCUS AF118569  
DEFINITION Homo sapiens angiotensin I converting enzyme precursor (Dcpl) gene,  
alternative splice products, complete cds.  
ACCESSION AF118569  
VERSION AF118569.1 GI:4732025  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS Rieder, M.J., Taylor, S.L., Clark, A.G. and Nickerson, D.A.  
TITLE Sequence variation in the human angiotensin converting enzyme  
JOURNAL Nat. Genet. 22 (1), 59-66 (1999)

MEDLINE 99251580  
PUBMED 10319862  
REFERENCE 2 (bases 1 to 24070)  
AUTHORS Rieder, M.J.  
TITLE Direct Submission  
JOURNAL Submitted (07-JAN-1999) Molecular Biotechnology, University of  
Washington, UW Health Sciences Building, Rm. K316, 1705 NE Pacific,  
Seattle, WA 98195, USA  
location/Qualifiers  
FEATURES  
source  
1. .24070  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17q23 between D17S794 and D17S795"  
117  
/frequency="0.05"  
/replace="L"  
334  
/frequency="0.09"  
/replace="L"  
complement(<337. .>488)  
/note="putative"  
/rpt\_family="L2"  
/rpt\_type-dispersed  
/note="putative"  
/rpt\_family="MIR"  
/rpt\_type-dispersed  
2400  
/frequency="0.32"  
/replace="L"  
2547  
/frequency="0.23"  
/replace="C"  
/gene="Dcpl"  
join(<2662. .2910,3498. .3665,4574. .4667,5336. .5479,  
5904. .6099,6658. .6755,7133. .7305,8032. .8255,8595. .8739,  
9026. .9124,9415. .9537,9896. .10107,10802. .10938,  
12123. .12281,12552. .12639,14499. .14642,14792. .14983,  
16805. .16902,17060. .17232,19287. .19510,19773. .19917,  
20223. .20321,22245. .22367,22649. .22836,22988. .>23217)  
/gene="Dcpl"  
/note="angiotensin I converting enzyme precursor"  
join(2662. .2910,3498. .3665,4574. .4667,5336. .5479,  
5904. .6099,6658. .6755,7133. .7305,8032. .8255,8595. .8739,  
9026. .9124,9415. .9537,9896. .10107,10802. .10938,  
12123. .12281,12552. .12639,14499. .14642,14792. .14983,  
16805. .16902,17060. .17232,19287. .19510,19773. .19917,  
20223. .20321,22245. .22367,22649. .22836,22988. .23217)  
/gene="Dcpl"  
/note="Somatic form of ACE; dipeptidyl carboxypeptidase 1;  
kininase II"  
/codon\_start=1  
/product="angiotensin I converting enzyme precursor"  
/protein\_id="AAD28560.1"  
/db\_xref="GI:4732026"  
/translation="MGAASGRGCGLLPPLLLPDPALADPGLQNGNSADAE  
GAQLFAOSYNSAEOVLFOSYAASMAHDNTITAEAROEBAALLSEFAWQOKAK  
ELYEPIWQNTFDPLRLRIGAVRTLGASNLPLARQOYNALLSMRSITYSTAKCLPN  
KTATCWSLDPDLTNIILASSRSYAMLFAWEGMNAAGIPLKPLYEDFTALSNAYKOD  
GFHTGAVRSWYNSPTFEDDLHLVQOLEPLYNLHAFVRAALHRRYGDYINLRGP  
IPALIGDMAOSWENIYDMVVPDPKPNLDVYSTMLQQGNATHMRVAEFTSLE  
LSPMPPEWEGSMLEKPADGREYVCHASANDFNKRPFRIKQCTRYMDOLSTVHNEM  
GHIOYLYQYKDLVSLRGANPGFHEAIGDLAVSTPEHLKIGLDRVTNDESD  
INYLKMALEKIAFLPPGYLVDMQRMVFGSGRPSPSNEEDMWYLRKYOGICPPVR  
NETHFDAGAKFHVNPYIRYFVSFVLQFHFALCKEYEGEPLHOCODYRSTKAG  
AKRKVLQAGSSRPMOEVLKDMVGLDADPLKYOPATYKOPATYKOPATYKOPATYKOP  
EYOWHPRLPNYPGIDLVNDEASKEFVEYEDTSOVVWNEVREANNVNTNTTET  
SKILLQKMOIANTHTLKYGOARKEFDYNQONTTKRIKKVODLERAPALPAOLEEY  
NKILLDMETTSVATVCHPNQSCQLDLEPLTNWATSRKEDILMAWEGMDKGRRI  
LOEPKYVELLINQARLVYDAGDSWRSWYEPSPSLDLERLFOELQPLYNLHAYV  
RRALHRYGAQHINLEGPILPAHLIGNMAQTSNITDLYVFPSPASMDTTEAMLKOG

WTPRRMFKREADDFETSLGLPPVPEPMNKSMLKPTDGEVYCHASANDFYNGKDFRI  
KOCITVNEIDEDVVAHHEKHIOIFMOYKDLVPLREBANGFHEAIGDVALSVSTPK  
HLHSLNLLSSGEGSDHDI NFMKMA LKIAI PPSYLVDMRWRFVDSITKENYQ  
EMWSSRLKYOGCLPPVPTGDPDGAGKFPSSVYIIEVFSIIQEFHIALCOAA  
GHTGPLHKCDIYOSKAGORLATAMKLGFRPMPEAMOLITGPPNMSASAMLSYFKPL  
LDMLRTEHGEKIGMPQYNWTPNSARSECP LPPSGRVSFGIDLDIDNOQAVGWML  
LFLGIALLVATIGLSORLFSIRHSIRHSHPOFGSEVELRHS"  
complement(2694..2737)  
repeat\_region  
/note="putative"  
/rpt\_type=tandem  
/rpt\_unit=cag  
3872  
variation  
/gene="DCP1"  
/frequency="0.32"  
/replace="a"  
/note="putative"  
/rpt\_family="Alu"  
/rpt\_type=dispersed  
4230  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="c"  
4504  
variation  
/gene="DCP1"  
/frequency="0.41"  
/replace="c"  
4733  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="a"  
4760  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="c"  
5406  
variation  
/gene="DCP1"  
/frequency="0.14"  
/replace="t"  
5424  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="a"  
5492  
variation  
/gene="DCP1"  
/frequency="0.09"  
/replace="t"  
5878  
variation  
/gene="DCP1"  
/frequency="0.09"  
/replace="c"  
6028  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="t"  
6029  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="t"  
6145  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="t"  
6604  
repeat\_region  
complement(6245..6320)  
/note="putative"  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
6435  
variation  
/gene="DCP1"  
/frequency="0.32"  
/replace="g"  
6604  
variation  
/gene="DCP1"

repeat\_region  
/frequency="0.05"  
/replace="a"  
/note="putative"  
/rpt\_family="Alu"  
/rpt\_type=dispersed  
7693  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="a"  
7831  
variation  
/gene="DCP1"  
/frequency="0.23"  
/replace="a"  
8128  
variation  
/gene="DCP1"  
/frequency="0.18"  
/replace="t"  
/note="putative"  
/rpt\_family="L2"  
/rpt\_type=dispersed  
8935  
repeat\_region  
/gene="DCP1"  
/frequency="0.05"  
/replace="c"  
8968  
variation  
/gene="DCP1"  
/frequency="0.36"  
/replace="t"  
9191  
variation  
/gene="DCP1"  
/frequency="0.09"  
/replace="t"  
9222  
variation  
/gene="DCP1"  
/frequency="0.09"  
/replace="t"  
9509  
variation  
/gene="DCP1"  
/frequency="0.05"  
/replace="c"  
8968  
variation  
/gene="DCP1"  
/frequency="0.36"  
/replace="t"  
9191  
variation  
/gene="DCP1"  
/frequency="0.09"  
/replace="t"  
9222  
variation  
/gene="DCP1"  
/frequency="0.09"  
/replace="t"  
9509  
variation  
/gene="DCP1"

Query Match 1.1%; Score 22; DB 9; Length 24070;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CAGAAGCCAGACCTGTATG 943  
DB 3543 CAGAAGCCAGACCTGTATG 3564

RESULT 30  
AC007360/c 80858 bp DNA linear PRI 30-SEP-2000  
LOCUS Homo sapiens BAC clone RP11-104K7 from 7q11.2-q22, complete  
DEFINITION  
ACCESSION AC007360  
VERSION AC007360.3 GI:10440739  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 80858)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 98063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 80858)  
AUTHORS Bourne, S., Holmes, A. and Maupin, R.  
TITLE The sequence of Homo sapiens BAC clone RP11-104K7  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 80858)



```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
misc.feature

```

Waterston,R.H.  
 Direct Submission  
 Submitted (23-APR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 4 (bases 1 to 80858)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (05-JUN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 5 (bases 1 to 80858)  
 Waterston,R.  
 Direct Submission  
 Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Sep 30, 2000 this sequence version replaced gi:5001526.  
 -----  
 Genome Center  
 -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 -----  
 Summary Statistics  
 -----  
 Center project name: H\_NH0104K07  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgrl.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgrl.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanesi,J.J. and de Jong,P.-J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBACe3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-356B17; the clone sequenced to the right is CTB-94H21, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-104K7; actual end is at base position 91620 of CTB-94H21.

Location/Qualifiers  
 1. 80858  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q11.2-q22"  
 /clone="RP11-104K7"  
 /clone\_lib="RPCI-11"  
 1. 350

```
misc.feature
repeat_region    1..336 /note="similar to EST A1360347 (NID:g4111968) qy89g04.xl  
                  /note="similar to EST T72671 (NID:g669346) yd19g06.sl"  
                  12..176 /rpt_family="Retroviral"  
repeat_region    350..401 /rpt_family="MER21"  
repeat_region    404..461 /rpt_family="L1"  
repeat_region    469..488 /rpt_family="L1"  
repeat_region    645..820 /rpt_family="(TAA)n"  
repeat_region    943..1185 /rpt_family="L1"  
repeat_region    1534..1765 /rpt_family="L1"  
repeat_region    1768..1856 /rpt_family="Alu"  
repeat_region    1875..2301 /rpt_family="L1"  
repeat_region    2684..2936 /rpt_family="L1"  
repeat_region    2937..3058 /rpt_family="MER4-group"  
repeat_region    3059..3357 /rpt_family="MER4-group"  
repeat_region    3358..3463 /rpt_family="Alu"  
repeat_region    3790..4000 /rpt_family="Alu"  
repeat_region    4404..4718 /rpt_family="L2"  
repeat_region    4753..4969 /rpt_family="Alu"  
repeat_region    5669..5716 /rpt_family="MIR"  
repeat_region    7163..7235 /rpt_family="MIR"  
repeat_region    7293..7331 /rpt_family="MIR"  
repeat_region    7686..7801 /rpt_family="(CA)n"  
repeat_region    8278..8321 /rpt_family="Alu"  
repeat_region    8322..8608 /rpt_family="MER21-group"  
repeat_region    8609..9272 /rpt_family="Alu"  
repeat_region    9402..9693 /rpt_family="MER21-group"  
repeat_region    9697..10140 /rpt_family="MER21-group"  
repeat_region    10161..10469 /rpt_family="L1"  
repeat_region    10497..10690 /rpt_family="Alu"  
repeat_region    10810..11053 /rpt_family="MER21-group"  
repeat_region    11056..11368 /rpt_family="MER4-group"  
repeat_region    11714..12015 /rpt_family="Alu"  
repeat_region    12384..12552 /rpt_family="Alu"  
repeat_region    12553..12849 /rpt_family="Alu"  
repeat_region    12850..12978 /rpt_family="Alu"  
repeat_region    13354..13553 /rpt_family="Alu"  
repeat_region    13603..13709 /rpt_family="MER4-group"  
repeat_region    /rpt_family="L1"
```

```

repeat_region 13710..14005 /rpt_family="Alu"
repeat_region 14006..14247 /rpt_family="L1"
repeat_region 14377..14536 /rpt_family="L1"
repeat_region 14599..14943 /rpt_family="MERL_type"
repeat_region 15193..15413 /rpt_family="MaLR"
repeat_region 15569..15616 /rpt_family="L2"
repeat_region 15972..16262 /rpt_family="MERL_type"
repeat_region 16469..16591 /rpt_family="Alu"
repeat_region 17610..17739 /rpt_family="Alu"
repeat_region 18087..18387 /rpt_family="Alu"
repeat_region 18641..18659 /rpt_family="L1"
repeat_region 18660..18958 /rpt_family="L1"
repeat_region 18959..19194 /rpt_family="Alu"
repeat_region 19220..19495 /rpt_family="L1"
repeat_region 19597..19725 /rpt_family="Alu"
repeat_region 19726..20052 /rpt_family="L1"
repeat_region 20055..20225 /rpt_family="MaLR"

```

```

Query Match 1.1%; Score 22; DB 9; Length 80858;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1366 AAAAAAGATTGCACCTTATG 1387
Db 71233 AAAAAAGATTGCACCTTATG 71212

```

```

RESULT 31
AC103946 201500 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP23-85E10, WORKING DRAFT SEQUENCE, 9 ordered
DEFINITION pieces.
AC103946 AC103946.3 GI:22381788
KEYWORDS HTG; HTGS_PHASe2; HTGS_DRAFT; HTGS_FULTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 201500)
AUTHORS Birten,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-85E10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201500)
AUTHORS Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gilde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McKean,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
COMMENT

```

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schpack,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,X., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 201500)  
 Birten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barina,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Chapel,Y., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., Fitzerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Linblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Rella,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schpack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thoman,N., Stojanovic,N., Talamas,J.,  
 Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced g1:19881886.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: 85\_R.10  
 Center clone name: 119635  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 199978 bases at least Q40  
 Consensus quality: 200513 bases at least Q30  
 Consensus quality: 200620 bases at least Q20  
 Insert size: 198000; agarose-fp  
 Insert size: 200700; sum-of-ctnigs  
 Quality coverage: 10.3 in Q20 bases; agarose-fp  
 Quality coverage: 10.2 in Q20 bases; sum-of-ctnigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 20114: contig of 20114 bp in length  
 \* 20115 20214: gap of 100 bp  
 \* 20215 24101: contig of 3887 bp in length  
 \* 24102 24201: gap of 100 bp  
 \* 24202 27267: contig of 3066 bp in length  
 \* 27268 27367: gap of 100 bp

```
* 27368 29341: contig of 1974 bp in length
* 29342 29441: gap of 100 bp
* 29442 33107: contig of 3666 bp in length
* 33108 33207: gap of 100 bp
* 33208 38389: contig of 5182 bp in length
* 38390 38489: gap of 100 bp
* 38490 129944: contig of 91455 bp in length
* 129945 130044: gap of 100 bp
* 130045 164792: contig of 34748 bp in length
* 164793 164892: gap of 100 bp
* 164893 201500: contig of 36608 bp in length.
FEATURES
    source
        1..201500
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone_1fb="RPCT-23 Female Mouse BAC"
            1..20114
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
            20215..24101
                /note="assembly_fragment"
            24202..27267
                /note="assembly_fragment"
            27368..29341
                /note="assembly_fragment"
            29442..33107
                /note="assembly_fragment"
            33208..38389
                /note="assembly_fragment"
            38490..129944
                /note="assembly_fragment"
            130045..164792
                /note="assembly_fragment"
            164893..201500
                /note="assembly_fragment"
                clone_end:T7
                vector_side:right"
BASE COUNT 60676 a 39450 c 40269 g 60305 t 800 others
ORIGIN
Query Match 1.1%; Score 22; DB 2; Length 201500;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2020 TCTTCTCCTTTGAAGTTGTCA 2041
    |||||
Db 43062 TCTTCTCCTTTGAAGTTGTCA 43083
RESULT 32
E50434
LOCUS E50434 21 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel protein.
ACCESSION E50434.1 GI:18633519
VERSION E50434.1 GI:18633519
KEYWORDS JP 2001057890-A/2.
SOURCE JP 2001057890-A/2.
ORGANISM JP 2001057890-A/2.
REFERENCE 1 (bases 1 to 21)
AUTHORS Omori,A. and Ito,M.
TITLE Novel protein
JOURNAL MITSUBISHI CHEM CORP
OS Artificial Sequence
PN JP 2001057890-A/2
PD 06-MAR-2001
PF 23-AUG-1999 JP 1999235218
PR
PI AKIRA OMORI, MAKOTO ITO
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/80, PC
```

```
C12N15/00, C12N5/00
CC
FH key Location/Qualifiers
FT source 1..21
FT /organism="Artificial Sequence".
FEATURES
    source
        1..21
            /organism="synthetic construct"
            /db_xref="taxon:32630"
BASE COUNT 6 a 1 c 8 g 6 t
ORIGIN
Query Match 1.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 562 CTGCTGATTGATGAGAA 582
    |||||
Db 1 CTGCTGATTGATGAGAA 21
RESULT 33
AF201331 4014 bp mRNA linear ROD 01-DEC-2000
LOCUS AF201331
DEFINITION Rattus norvegicus strain F344/N angiotensin-converting enzyme (Ace)
ACCESSION AF201331.1 GI:11493660
VERSION AF201331.1
KEYWORDS
SOURCE AF201331.1
ORGANISM Rattus norvegicus.
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 4014)
Jafarian-Tehrani,M., Listwak,S., Barrientos,R.M., Michaud,A.,
Corvol,P. and Sternberg,E.M.
Characterization of a missense mutation in the angiotensin
I-converting enzyme cDNA in exudative inflammation resistant F344/N
rats
Unpublished
2 (bases 1 to 4014)
Jafarian-Tehrani,M., Listwak,S., Barrientos,R.M., Michaud,A.,
Corvol,P. and Sternberg,E.M.
Direct Submission
Submitted (02-NOV-1999) CNE, NIMH, 10 Center Dr., Bethesda, MD
20892-1284, USA
FEATURES
    source
        1..4014
            Location/Qualifiers
            1..4014
                /organism="Rattus norvegicus"
                /strain="F344/N"
                /db_xref="taxon:10116"
                /chromosome="10"
                /tissue_type="lung"
            1..4014
                /gene="Ace"
            1..23
                /gene="Ace"
            24..3965
                /gene="Ace"
            /codon_start=1
            /product="angiotensin-converting enzyme"
            /protein_id="AA035596.1"
            /db_xref="GI:11493661"
            /translation="MGASGQGRWPLSPPLMLSLLLLLPPSPAPALDGLQPN
            FSADGAGQLPADSYNSAEVYFQSTASVAHPTNITEENARQEEALINQFAEV
            KGRKAKELYESIMQWFTDQKLRIIGSVOTGPNLPLTORLQVNSLNSRSRYSTG
            KVCPEPKATGCSIDPELITLASRNYAKVLFAMEGWDVAGIPIRLYDDFTALSN
            EAYRQDGSDDTGATYRSYTESPESIEHXYHOVEPYLTHAFYRRLARRGDKY
            INLGPILPAHLGDMQASQWENITDMVVPFDDKPLDVTSTVQKGNATHTFVAEE
            FETSLGSPMPPEEEMSLERKPADREVVCASAMDEYRNRKRIKQCTATVTDQUS
            TVHHEMGVQYLYQKDLHVSILRGANGFEALIGDVALSVSTPAHLHKIGLIDRYA
            NDIESDINYLKMLEKIAFLPFGYLVDQRMGVFSGRTPPSRYNIDMWYLRTRYGI
```



```

5/UTR      1..23
CDS        /gene="ACE"
           24..3965
           /gene="ACE"
           /codon_start=1
           /product="angiotensin converting enzyme"
           /protein_id="AA82110.1"
           /db_xref="GI:437288"
           /translation="MGAASGQGRWMLSPPLMLSLLLLLLPPSPAPALDPGLQPGN
FSADAEQIQLFADSYNSSAEYVQFSQTAASGAHDTNTITEARALDQFEAEYAEYAEY
WGKAKLYESIQWQNFQDKLRIGISVYATLQFPELOYNLSLNSNRISYTGSTG
KVCPPNKTATCWSLDPETLNTLASRNNAKFLFMWGHDDVGIPLRLPYODFTALS
EARRQDFSDTGAAYWRWSWESPSSESLHHYHVEPLYLHAFVRALRRIGDKK
INLRGPIPAHLGDMMQSMENITDMVVPEDKRLDVTSMVGKWTATMFRAEE
EFTSLSPMPPEFWMAESMLERPADGRBVCVHASAMFQYNKDRIKQCTVTMDQLS
TVHEHSHVQYLYQKDLHVSILRGANGDFEALGDVLTALSVSPNHLKILGLDRVAT
NDIESDINYLILKMLLEKIAFLPEGYLVDMQWMEVSGEPTSPRNHMKYLRTRYOG
CPYARNEHTHDGAKRHPISVPYTRFVSVFVQFOPHQLCKEAGHSGTLHOCDTIY
QSTRAGAKLDQVLDAGCSRPMQEVKLKLVSDALDASHMLTYFQVPSQWLEQNRQNK
EVLGAEYQMRPLPDVYPEGIDLETDEANRVEEEDRYAKLAMEYALQANMYN
NITIEGSKILLQKKEVSNHLLKYGTAKEFDENFQNSPTIKRIKKQVNDRAVLR
NELIEYQIILLDMETTSYVAVNYCTNTGSLDELDTNIMTISRKEVTLGWSMSR
KACNADWMSLRLKLYOGICLPVPSQGDPDFGSFHPANPYLRITLRFSLTIQRFMSA
ALCRAAGHTGPIKCDIYQSGEAGKLLADAMKLGYSKQMPAMKTIIGQPMASAIITK
NYRRPLLEMLYTERRNGETLGMPEYTWTPNTARAESLPESSRVNLTGMTLEFQQA
VGVQVLLFGLVALVATVGLAHLRLYNHNHSLRRPHRGPGSGSEVELRHS"
3/UTR      3966..4014
           /gene="ACE"
BASE COUNT 961 a 1150 c 1080 g 823 t
ORIGIN

Query Match      1.0%; Score 21; DB 10; Length 4014;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      923 AGAAGGCCAAGAGCTGATG 943
Db      337 AGAAGGCCAAGAGCTGATG 357

RESULT 36
RND03734      4142 bp      linear      ROD 30-NOV-1995
LOCUS      Rattus norvegicus Wistar-Kyoto (Heidelberg) angiotensin converting
DEFINITION      enzyme (ACE) mRNA, complete cds.
ACCESSION      U03734
KEYWORDS
SOURCE      U03734.1 GI:437289
ORGANISM      Rat.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE      1 (bases 1 to 4142)
AUTHORS      Koike,G., Krieger,J.E., Jacob,H.J., Mukoyama,M., Pratt,R.E. and
            Ddau,V.J.
TITLE      Angiotensin converting enzyme and genetic hypertension: cloning of
            rat cDNAs and characterization of the enzyme
JOURNAL      Biochem. Res. Commun. 198 (1), 380-386 (1994)
MEDLINE      94121658
PubMed      8292044
FEATURES
            2 (bases 1 to 4142)
            Direct Submission
            Koike,G.
            Submitted (26-NOV-1993) George Koike, Falk Cardiovascular Research
            Center, Stanford University, School of Medicine, 300 Pasteur Drive,
            Palo Alto, CA 94305-5246, USA
            Location/Qualifiers
FEATURES

```

Source	1. .4142	/organism="Rattus norvegicus"	/strain="Wistar-Kyoto (Heidelberg); WKY"	/db_xref="taxon:10116"	/clone="pGRW154, pGRW102"	/tissue_type="lung"	/clone_lib="Lambda zap II cDNA library"	gene	1. .4142	/gene="ACE"	1. .23	/gene="ACE"	24. .3965	/gene="ACE"	/codon_start=1	/product="angiotensin converting enzyme"	/protein_id="AA8211.1"	/db_xref="GI:437290"	/translation="MGAAAGGGRWPLSPPLMLSLILLPLPPSPADLPGLPGNGKRAKELYESIWONFTDQKLRIRIGSVOTLGRNLLPRLQRLNLSLSMSRITSTGKVCPCNTATCWASLDPELNTNLASSRNYAVLFWMECHDAVGPILKPLVODFPALSN EAYQDQSDTGAWRSWYBSPSEESLEHLHYOPELYNLHAFVRRALRRYGDYK IINLSPGIPALHLDGMAQSMENIYDMVPEPPCKNLIDVTSTWQKGNATIMFVAAEE PFNSLSPMPPEFMAASMLEKPADGRRVCHASAMPVFNKDRICQCRPTVMDLS TVHMEKHVOYILOYKDLVHSLRGANGPHEHAGDVALSVSPAILHKIGLDRVPA NDIESLNTLTKMALEIAFLPEGYLDVDMGWGVSGRTPPSRTNYGMWYLIRKVOGI CIPARNEITHEDAGAKFHIPISVPIRYFVSPVLOQFPHALCKEAGHOGFLHOCIDIT YOSTGAKLQOVLQAGCSRPWOEVLKLVGSDALDASALMEYEPQVQMLOEONRNGN ECTGAPKLOOVPPLPDNPPEGIDLETDEKANRVEYEDIRAKVLMVEYEAANHNTYR NITREGSKILQKKKYSNHTIKYGTAKTFEDVSNPNQSTRIKIKKQONDAVLEPP NELEFQOILDMETTYTAVKANVCYTNCTGSLSEPLDNLNATSKRYELLNWKSPMD KVGAILLPEPPKYVDVFNSTKLAKLNGYSIADGDSVSESDLEODLEKLVOELPVLV LNHAYVRSLSRHNGSEYINLDGPIPALHLDGMAQSMENIYDMVPEPPSPASIDATE ALAIQGTGPRRIEKEADNFTSLDILPVLPEGMFKSMLEKTDGREVCHASADPFYN GKDRIKQCTSVNNEELVIAHMEGHIOYFMQYKDLVTFREGANPGEHAGDVALA SVSPKRLHSINILSSRSGYEHINFLMKMLDKIAFLPEPSYLDQMRMRVFGSITTS KENTNQMWSRLRKQGLCPRPVRSQSDPDGSGFTHPANVPIRITRISTLIQPHHNE ALCAAGHTGPLYKCDIYOSKEAGKLADAKLQYSKQWPEAMKITTGQPMWSASAIM NYKFLPEMLVETENRRHGETLGMPETYMTPTNARAESESSRVNLTGMVLEPOOAR VGVQVLLFLGVALLVATVGLHRLRYLNHNHSLRRPRGQFGSEVELRIS"	3'UTR	BASE COUNT	984 a	1194 c	1106 g	858 t	ORIGIN	Query Match	Best Local Similarity	100.0%	Score 21;	DB 10;	Length 4142;	LOCUS	AL591419/c	DEFINITION	Human DNA sequence from clone Rpl1-134L4 on chromosome 6, complete sequence.	ACCESSION	AL591419	VERSION	AL591419.2	GI:14422240	SOURCE	Human.	ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	REFERENCE	1 (bases 1 to 53423)	Laïrd, G.	Direct Submission	Submitted (11-JUN-2001)	Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerey@sanger.ac.uk	Requests: clone requests@sanger.ac.uk	On Jun 13, 2001 this sequence version replaced gi:14133300.	COMMENT
gene	1. .4142	/organism="Rattus norvegicus"	/strain="Wistar-Kyoto (Heidelberg); WKY"	/db_xref="taxon:10116"	/clone="pGRW154, pGRW102"	/tissue_type="lung"	/clone_lib="Lambda zap II cDNA library"	gene	1. .4142	/gene="ACE"	1. .23	/gene="ACE"	24. .3965	/gene="ACE"	/codon_start=1	/product="angiotensin converting enzyme"	/protein_id="AA8211.1"	/db_xref="GI:437290"	/translation="MGAAAGGGRWPLSPPLMLSLILLPLPPSPADLPGLPGNGKRAKELYESIWONFTDQKLRIRIGSVOTLGRNLLPRLQRLNLSLSMSRITSTGKVCPCNTATCWASLDPELNTNLASSRNYAVLFWMECHDAVGPILKPLVODFPALSN EAYQDQSDTGAWRSWYBSPSEESLEHLHYOPELYNLHAFVRRALRRYGDYK IINLSPGIPALHLDGMAQSMENIYDMVPEPPCKNLIDVTSTWQKGNATIMFVAAEE PFNSLSPMPPEFMAASMLEKPADGRRVCHASAMPVFNKDRICQCRPTVMDLS TVHMEKHVOYILOYKDLVHSLRGANGPHEHAGDVALSVSPAILHKIGLDRVPA NDIESLNTLTKMALEIAFLPEGYLDVDMGWGVSGRTPPSRTNYGMWYLIRKVOGI CIPARNEITHEDAGAKFHIPISVPIRYFVSPVLOQFPHALCKEAGHOGFLHOCIDIT YOSTGAKLQOVLQAGCSRPWOEVLKLVGSDALDASALMEYEPQVQMLOEONRNGN ECTGAPKLOOVPPLPDNPPEGIDLETDEKANRVEYEDIRAKVLMVEYEAANHNTYR NITREGSKILQKKKYSNHTIKYGTAKTFEDVSNPNQSTRIKIKKQONDAVLEPP NELEFQOILDMETTYTAVKANVCYTNCTGSLSEPLDNLNATSKRYELLNWKSPMD KVGAILLPEPPKYVDVFNSTKLAKLNGYSIADGDSVSESDLEODLEKLVOELPVLV LNHAYVRSLSRHNGSEYINLDGPIPALHLDGMAQSMENIYDMVPEPPSPASIDATE ALAIQGTGPRRIEKEADNFTSLDILPVLPEGMFKSMLEKTDGREVCHASADPFYN GKDRIKQCTSVNNEELVIAHMEGHIOYFMQYKDLVTFREGANPGEHAGDVALA SVSPKRLHSINILSSRSGYEHINFLMKMLDKIAFLPEPSYLDQMRMRVFGSITTS KENTNQMWSRLRKQGLCPRPVRSQSDPDGSGFTHPANVPIRITRISTLIQPHHNE ALCAAGHTGPLYKCDIYOSKEAGKLADAKLQYSKQWPEAMKITTGQPMWSASAIM NYKFLPEMLVETENRRHGETLGMPETYMTPTNARAESESSRVNLTGMVLEPOOAR VGVQVLLFLGVALLVATVGLHRLRYLNHNHSLRRPRGQFGSEVELRIS"	3'UTR	BASE COUNT	984 a	1194 c	1106 g	858 t	ORIGIN	Query Match	Best Local Similarity	100.0%	Score 21;	DB 10;	Length 4142;	LOCUS	AL591419/c	DEFINITION	Human DNA sequence from clone Rpl1-134L4 on chromosome 6, complete sequence.	ACCESSION	AL591419	VERSION	AL591419.2	GI:14422240	SOURCE	Human.	ORGANISM	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	REFERENCE	1 (bases 1 to 53423)	Laïrd, G.	Direct Submission	Submitted (11-JUN-2001)	Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerey@sanger.ac.uk	Requests: clone requests@sanger.ac.uk	On Jun 13, 2001 this sequence version replaced gi:14133300.	COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, SW: SWISSPROT, Tr: TrEMBL, Mp: MOPREP; information on the MOPREP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/moprep](http://www.sanger.ac.uk/Projects/C_elegans/moprep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/MGP/Chr6>

RP11-134L4 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-134L4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-134L4 is at 53423 in this sequence. The true right end of clone RP11-486M3 is at 100 in this sequence.

## FEATURES

## Source

```
1. 53423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-134L4"
/clone_id="RPCI-11.1"
/complement(1..114)
/note="match: GSS: Em:AQ132196"
/complement(1..94)
/note="match: GSS: Em:AQ277160"
304..612
/note="match: GSS: Em:AQ277160"
/note="Alusg repeat: matches 1..308 of consensus"
756..947
/note="MER63A repeat: matches 5..202 of consensus"
1148..1243
/note="48 copies 2 mer tt 61% conserved"
1262..1458
/note="match: GSS: Em:AQ081950"
1344..1485
/note="FLAM_C repeat: matches 2..140 of consensus"
1490..1809
/note="Alusx repeat: matches 1..312 of consensus"
2477..3001
/note="L2 repeat: matches 2278..2735 of consensus"
3287..3477
/note="match: GSS: Em:B95534"
3482..3650
/note="match: GSS: Em:B95534"
3786..4147
/note="match: GSS: Em:AQ672311"
4156..4460
/note="Alusp repeat: matches 1..307 of consensus"
4588..4778
/note="MIR repeat: matches 2..208 of consensus"
5837..5991
/note="MIR repeat: matches 34..185 of consensus"
6610..6736
/note="L2 repeat: matches 2619..2746 of consensus"
7387..7785
/note="L1MC repeat: matches 272..671 of consensus"
```

```
repeat_region 7812..9031
/note="L1MC repeat: matches 1331..2435 of consensus"
repeat_region 9032..9342
/note="Alusx repeat: matches 1..309 of consensus"
repeat_region 9343..9656
/note="L1MC repeat: matches 2435..2741 of consensus"
misc_feature 10822..13811
/complement(10314..10795)
/note="match: GSS: Em:AQ835745"
repeat_region 14336..14421
/note="L1M4 repeat: matches 272..2803 of consensus"
repeat_region 14347..14835
/note="MIR repeat: matches 63..149 of consensus"
misc_feature 15243..15683
/note="Aluy repeat: matches 1..289 of consensus"
/complement(14845..15241)
/note="match: GSS: Em:AQ221492"
15243..15683
/note="match: GSS: Em:AQ296090"
16434..16742
/note="Alusg repeat: matches 1..308 of consensus"
18333..18395
/note="L2 repeat: matches 2668..2731 of consensus"
18609..18842
/note="Alub repeat: matches 1..234 of consensus"
18927..19219
/note="Alub repeat: matches 1..292 of consensus"
19334..19673
/note="L2 repeat: matches 2224..2356 of consensus"
19890..20426
/note="match: GSS: Em:AQ081102"
19890..20412
/note="match: GSS: Em:AQ081070"
19890..20330
/note="match: GSS: Em:B88888"
19932..20397
/note="match: GSS: Em:AQ014577"
20610..21212
/note="match: GSS: Em:AQ787006"
20888..21119
/note="Alusx repeat: matches 1..228 of consensus"
21150..21435
/note="Alusx repeat: matches 1..300 of consensus"
21336..21600
/note="Alusg/x repeat: matches 134..297 of consensus"
21823..22102
/note="Aluo repeat: matches 1..282 of consensus"
22115..22306
/note="L1MA10 repeat: matches 5981..6322 of consensus"
24101..24483
/note="WSTB repeat: matches 1..426 of consensus"
/complement(24849..25332)
25497..25665
/note="match: GSS: Em:AQ77587"
25517..26616
/note="L2 repeat: matches 2319..2488 of consensus"
26617..26920
/note="L2 repeat: matches 2500..2605 of consensus"
26921..27079
/note="Alusx repeat: matches 1..304 of consensus"
27095..27181
/note="L2 repeat: matches 2605..2735 of consensus"
27774..27865
/note="L2 repeat: matches 2656..2745 of consensus"
27941..28237
/note="46 copies 2 mer ca 69% conserved"
28385..28679
/note="Alusx repeat: matches 1..298 of consensus"
28875..29019
/note="Alub repeat: matches 1..284 of consensus"
30363..30458
/note="MIR repeat: matches 48..195 of consensus"
31215..31457
/note="HERV1A0 repeat: matches 4..107 of consensus"
```



2205 2304: gap of unknown length  
2305 3320: contig of 1016 bp in length  
3321 3420: gap of unknown length  
3421 4474: contig of 1054 bp in length  
4475 5704: gap of unknown length  
5705 5804: contig of 1130 bp in length  
5805 7210: gap of unknown length  
7211 8442: contig of 1132 bp in length  
8443 8542: gap of unknown length  
8543 10195: contig of 1653 bp in length  
10196 10285: gap of unknown length  
10286 11430: contig of 1195 bp in length  
11431 11590: gap of unknown length  
11591 13046: contig of 1456 bp in length  
13047 13146: gap of unknown length  
13147 14505: contig of 1359 bp in length  
14506 14605: gap of unknown length  
14606 16138: contig of 1533 bp in length  
16139 17882: contig of 1544 bp in length  
17883 17882: gap of unknown length  
17883 19585: contig of 1713 bp in length  
19586 19685: gap of unknown length  
19686 20790: contig of 1095 bp in length  
20791 20890: gap of unknown length  
20891 22384: contig of 1494 bp in length  
22385 22484: gap of unknown length  
22485 23511: contig of 1027 bp in length  
23512 23611: gap of unknown length  
23612 25171: contig of 1560 bp in length  
25172 25271: gap of unknown length  
25272 27046: contig of 1775 bp in length  
27047 27146: gap of unknown length  
27147 28748: contig of 1602 bp in length  
28749 28848: gap of unknown length  
28849 29995: contig of 1147 bp in length  
29996 30095: gap of unknown length  
30096 31696: contig of 1601 bp in length  
31697 32947: contig of 1151 bp in length  
32948 33047: gap of unknown length  
33048 34144: contig of 1097 bp in length  
34145 34244: gap of unknown length  
34245 35729: contig of 1485 bp in length  
35730 35829: gap of unknown length  
35830 37855: contig of 2026 bp in length  
37856 37955: gap of unknown length  
37956 39379: contig of 1424 bp in length  
39380 39479: gap of unknown length  
39480 41067: contig of 1588 bp in length  
41068 41167: gap of unknown length  
41169 42304: contig of 1137 bp in length  
42305 42404: gap of unknown length  
42406 44115: contig of 1711 bp in length  
44116 44215: gap of unknown length  
44217 45401: contig of 1186 bp in length  
45402 45501: gap of unknown length  
45502 47109: contig of 1608 bp in length  
47110 47209: gap of unknown length  
47211 50085: contig of 2876 bp in length  
50086 50185: gap of unknown length  
50186 53328: contig of 3143 bp in length  
53329 53428: gap of unknown length  
53429 55854: contig of 2426 bp in length  
55855 55954: gap of unknown length  
55955 58755: contig of 2801 bp in length  
58756 58855: gap of unknown length  
58856 60750: contig of 1895 bp in length  
60751 60850: gap of unknown length  
60851 62682: contig of 1833 bp in length  
62682 62782: gap of unknown length

62783 64194: contig of 1412 bp in length  
64195 64294: gap of unknown length  
64295 65949: contig of 1655 bp in length  
65950 66049: gap of unknown length  
66050 68333: contig of 2284 bp in length  
68334 68433: gap of unknown length  
68434 70841: contig of 2408 bp in length  
70842 70941: gap of unknown length  
70942 73306: contig of 2365 bp in length  
73307 73406: gap of unknown length  
73407 76700: contig of 3294 bp in length.  
Location/Qualifiers  
1..76700  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-193C19"  
BASE COUNT 20693 a 15464 c 14004 g 21005 t 5534 others  
ORIGIN  
Query Match 1.0%; Score 21; DB 2: Length 76700;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2021 CTCCTCCTTTTGAAGTTGCA 2041  
Db 14318 CTCCTCCTTTTGAAGTTGCA 14338  
RESULT 39  
AC096435  
LOCUS  
DEFINITION  
AC096435 Rattus norvegicus clone CH230-44G5. \*\*\* SEQUENCING IN PROGRESS \*\*\*  
60 unordered pieces.  
AC096435  
AC096435.3 GI:21723831  
VERSION  
KEYWORDS  
HTG, HTGS, PHASE1.  
SOURCE  
ORGANISM  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 100858)  
REFERENCE  
AUTHORS  
Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbarta,J., Benton,J., Blimege,K., Blankenbury,K., Bonnin,D.,  
Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Doulhwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,F., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homs,F., Howard,S., Huber,J., Huiyk,S., Hune,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssohn,E., Kelly,S., Khan,U., King,U., Korvan,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,N., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokkenko,S., Oguh,M., Okunolu,G.,  
Oraguwe,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Picken,R., Pimus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G.,





```

* 70605 71873: contig of 1269 bp in length
* 71874 71973: gap of unknown length
* 71974 73634: contig of 1661 bp in length
* 73635 73734: gap of unknown length
* 73735 75766: contig of 2032 bp in length
* 75767 75866: gap of unknown length
* 75867 78639: contig of 2773 bp in length
* 78640 78739: gap of unknown length
* 78740 80125: contig of 1386 bp in length
* 80126 80225: gap of unknown length
* 80226 82615: contig of 2390 bp in length

```

Query Match 1.0%; Score 21; DB 2; Length 100858;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 923 AGAGGCCAAGGAGCTGTATG 943  
 Db 3188 AGAGGCCAAGGAGCTGTATG 3208

```

RESULT 40
AC098290/c
LOCUS
DEFINITION
AC098290
VERSION
AC098290.3 GI:21953793
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 179743)
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaratunge,H.C., Are,J.R., Ayalew,M., Banks,T.,
Bairdita,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,D., Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Holling,B.,
Homai,F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kraevic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozdo,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Plinius,E., Pu,L.L., Qulles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubenok,I., Rolfe,R., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tatney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,

```

```

TITLE
JOURNAL
Unpublished
2 (bases 1 to 179743)
AUTHORS
Worley,K.C.
DIRECT SUBMISSION
Submitted (23-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 179743)
REFERENCE
Worley,K.C.
DIRECT SUBMISSION
Submitted (31-Jul-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:20976531.
COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUGA
Center clone name: CH230-1K15
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 13139 bases at least Q40
Consensus quality: 137136 bases at least Q30
Consensus quality: 141601 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 72 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1138: contig of 1138 bp in length
1139 1238: gap of unknown length
1239 2301: contig of 1063 bp in length
2302 2401: gap of unknown length
2402 3496: contig of 1095 bp in length
3497 3596: gap of unknown length
3597 5069: contig of 1473 bp in length
5070 5169: gap of unknown length
5169 6811: contig of 1642 bp in length
6812 8070: contig of 1159 bp in length
8071 8170: gap of unknown length
8171 9289: contig of 1129 bp in length
9290 9399: gap of unknown length
9399 10818: contig of 1419 bp in length
10819 12098: gap of unknown length
12099 12195: contig of 1177 bp in length
12196 13277: gap of unknown length
13278 13377: gap of unknown length
13377 14438: contig of 1061 bp in length
14439 14538: gap of unknown length
14539 16006: contig of 1468 bp in length
16007 16107: gap of unknown length
16107 18128: contig of 2022 bp in length
18129 18228: gap of unknown length
18229 19676: contig of 1448 bp in length
19677 21436: gap of unknown length
21437 21536: gap of unknown length

```

```

* 21537 22548: contig of 1012 bp in length
* 22549 22648: gap of unknown length
* 22649 24318: contig of 1670 bp in length
* 24319 24418: gap of unknown length
* 24419 25580: contig of 1162 bp in length
* 25581 25680: gap of unknown length
* 25681 27093: contig of 1413 bp in length
* 27094 27193: gap of unknown length
* 27194 28470: contig of 1277 bp in length
* 28471 28570: gap of unknown length
* 28571 30684: contig of 2114 bp in length
* 30685 30784: gap of unknown length
* 30785 31851: contig of 1067 bp in length
* 31852 31951: gap of unknown length
* 31952 33328: contig of 1377 bp in length
* 33329 33428: gap of unknown length
* 33429 35336: contig of 1908 bp in length
* 35337 35436: gap of unknown length
* 35437 37015: contig of 1579 bp in length
* 37016 37115: gap of unknown length
* 37116 38430: contig of 1315 bp in length
* 38431 38530: gap of unknown length
* 38531 40403: contig of 1873 bp in length
* 40404 40503: gap of unknown length
* 40504 41571: contig of 1068 bp in length
* 41572 41671: gap of unknown length
* 41672 43207: contig of 1536 bp in length
* 43208 43307: gap of unknown length
* 43308 43175: contig of 1868 bp in length
* 43176 45275: gap of unknown length
* 45276 47055: contig of 1780 bp in length
* 47056 47155: gap of unknown length
* 47156 49525: contig of 2370 bp in length
* 49526 49625: gap of unknown length
* 49626 51279: contig of 1654 bp in length
* 51280 51379: gap of unknown length
* 51380 53441: contig of 2062 bp in length
* 53442 53541: gap of unknown length
* 53542 55028: contig of 1487 bp in length
* 55029 55128: gap of unknown length
* 55129 56806: contig of 1678 bp in length
* 56807 56906: gap of unknown length
* 56907 58557: contig of 1751 bp in length
* 58558 58758: gap of unknown length
* 58759 60008: contig of 1251 bp in length
* 60009 60108: gap of unknown length
* 60109 62324: contig of 2216 bp in length
* 62325 62424: gap of unknown length
* 62425 64265: contig of 1841 bp in length
* 64266 64365: gap of unknown length
* 64366 65759: contig of 1394 bp in length
* 65760 65859: gap of unknown length
* 65860 67579: contig of 1720 bp in length
* 67580 67679: gap of unknown length
* 67680 69047: contig of 1368 bp in length
* 69048 71865: gap of unknown length
* 71866 71965: contig of 2718 bp in length
* 71966 75130: contig of 3165 bp in length
* 75131 75230: gap of unknown length
* 75231 78455: contig of 3225 bp in length
* 78456 78555: gap of unknown length
* 78556 80136: contig of 1581 bp in length
* 80137 80236: gap of unknown length
* 80237 82812: contig of 2576 bp in length
* 82813 82912: gap of unknown length
* 82913 84864: contig of 1952 bp in length
* 84865 84964: gap of unknown length
* 84965 87080: contig of 2116 bp in length
* 87081 87180: gap of unknown length
* 87181 89215: contig of 2035 bp in length
* 89216 89315: gap of unknown length
* 89316 92002: contig of 2687 bp in length

```

```

* 92003 92102: gap of unknown length
* 92103 94959: contig of 2857 bp in length
* 94960 95059: gap of unknown length
* 95060 97782: contig of 2723 bp in length

Query Match 1.0%; Score 21; DB 2; Length 179743;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1220 TGCTTCACAGCGAGAGCTGA 1240
Db 44323 TGCTTCACAGCGAGAGCTGA 44303

RESULT 41
AC068982
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC068982 181663 bp DNA linear HTG 07-JUL-2000
Homo sapiens chromosome 6 clone RP11-79506, WORKING DRAFT SEQUENCE,
7 unordered pieces.
AC068982
AC068982.3 GI:8099316
HTG, HTGS_PHASE1, HTGS_DRAFT.
Homo sapiens.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181663)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 181663)
Waterston, R.H.
Direct Submission
Submitted (15-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 27, 2000 this sequence version replaced gi:7924013.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0795006
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: Plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175821 bases at least Q40
Consensus quality: 177817 bases at least Q30
Consensus quality: 178903 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 181063; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 6.15 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2688: contig of 2688 bp in length
* 2689 2788: gap of unknown length
* 2789 9166: contig of 6378 bp in length
* 9167 9266: gap of unknown length
* 9267 17646: contig of 8380 bp in length
* 17647 17746: gap of unknown length
* 17747 32787: contig of 15041 bp in length

```

```

* 32788 32887: gap of unknown length
* 32888 49393: contig of 16506 bp in length
* 49394 49493: gap of unknown length
* 49494 99672: contig of 50179 bp in length
* 99673 99772: gap of unknown length
* 99773 181663: contig of 81881 bp in length.
FEATURES
    source
        1. 181663
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="RP11-79506"
            1. 2688
            /note="assembly_name:Contig8"
            2789. 9166
            /note="assembly_name:Contig9"
            clone_end:17
            vector_side:left"
            9267. 17646
            /note="assembly_name:Contig10"
            17747. 32787
            /note="assembly_name:Contig11"
            32888. 49393
            /note="assembly_name:Contig12"
            49494. 99672
            /note="assembly_name:Contig13"
            99773. 181663
            /note="assembly_name:Contig14"
BASE COUNT  52019 a 37585 c 38060 g 53390 t      609 others
ORIGIN
Query Match      1.0%; Score 21; DB 2; Length 181663;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1358 AGCGAATTAAAAAGATTG 1378
Db 116879 AGCGAATTAAAAAGATTG 116899

RESULT 42
LOCUS AC120772/c
DEFINITION Rattus norvegicus clone CH230-203F10, *** SEQUENCING IN PROGRESS
ACCESSION AC120772
VERSION AC120772.3 GI:21909159
KEYWORDS HTG: HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
          1 (bases 1 to 186651)
          Muszty,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
          Alstropoos,S.L., Amaralungu,H.C., Are,J.R., Ayale,M., Banks,T.,
          Barabara,J., Benton,J., Blimberg,K., Blankenburg,K., Bonnin,D.,
          Bouck,J., Bowles,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
          Buhay,C., Burch,P., Burkett,C., Burrell,R.L., Byrd,K.C.,
          Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
          Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
          Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Doukwalte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Giorrelli,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
          Hernandez,O., Hodson,A., Hogues,M., Holloway,C., Hollins,B.,
          Homel,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
          Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Karlssohn,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovoll,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Maxwell,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
Nguyen,N., Nickerson,E., Nwokenko,S., Ogih,M., Okunou,G.,
Oradunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoan,T., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,T.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Wainstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 186651)
Worley,K.C.
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 186651)
Worley,K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 19, 2002 this sequence version replaced gi:20564394.
-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----Project Information
Center project name: GYCS
Center clone name: CH230-203F10
-----Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 142024 bases at least Q40
Consensus quality: 148489 bases at least Q30
Consensus quality: 152361 bases at least Q20
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a working draft sequence. It currently
consists of 63 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1113: contig of 1113 bp in length
* 1114 1213: gap of unknown length
* 1214 2488: contig of 1275 bp in length
* 2489 2588: gap of unknown length
* 2589 4057: contig of 1469 bp in length
* 4058 4157: gap of unknown length
* 4158 5351: contig of 1194 bp in length
* 5352 7014: gap of unknown length
* 7015 7114: contig of 1563 bp in length
* 7115 8812: contig of 1698 bp in length

```

```

*      8813      8912: gap of unknown length
*      8913      9958: contig of 1046 bp in length
*      9959      10058: gap of unknown length
*     10059      11113: contig of 1055 bp in length
*     11114      11213: gap of unknown length
*     11214      12318: contig of 1105 bp in length
*     12319      12418: gap of unknown length
*     12419      14029: contig of 1611 bp in length
*     14030      14129: gap of unknown length
*     14130      15613: contig of 1484 bp in length
*     15614      15713: gap of unknown length
*     15714      16851: contig of 1138 bp in length
*     16852      16951: gap of unknown length
*     16952      18585: contig of 1634 bp in length
*     18586      18685: gap of unknown length
*     18686      20981: contig of 2296 bp in length
*     20982      21081: gap of unknown length
*     21082      23190: contig of 2109 bp in length
*     23191      23290: gap of unknown length
*     23291      24804: contig of 1514 bp in length
*     24805      24904: gap of unknown length
*     24905      26496: contig of 1592 bp in length
*     26497      26596: gap of unknown length
*     26597      28532: contig of 1936 bp in length
*     28533      28632: gap of unknown length
*     28633      30042: contig of 1410 bp in length
*     30043      30142: gap of unknown length
*     30143      31313: contig of 1171 bp in length
*     31314      31413: gap of unknown length
*     31414      32881: contig of 1468 bp in length
*     32882      32981: gap of unknown length
*     32982      34460: contig of 1479 bp in length
*     34461      34560: gap of unknown length
*     34561      36685: contig of 2125 bp in length
*     36686      36785: gap of unknown length
*     36786      38279: contig of 1494 bp in length
*     38280      38379: gap of unknown length
*     38380      39958: contig of 1579 bp in length
*     39959      40058: gap of unknown length
*     40059      41767: contig of 1709 bp in length
*     41768      41867: gap of unknown length
*     41868      44854: contig of 2987 bp in length
*     44855      44954: gap of unknown length
*     44955      46927: contig of 1973 bp in length
*     46928      47027: gap of unknown length
*     47028      48811: contig of 1784 bp in length
*     48812      48911: gap of unknown length
*     48912      50907: contig of 1996 bp in length
*     50908      51007: gap of unknown length
*     51008      54318: contig of 3311 bp in length
*     54319      54418: gap of unknown length
*     54419      55922: contig of 1504 bp in length
*     55923      56022: gap of unknown length
*     56023      58319: contig of 2297 bp in length
*     58320      58419: gap of unknown length
*     58420      60899: contig of 2480 bp in length
*     60900      60999: gap of unknown length
*     61000      63820: contig of 2821 bp in length
*     63821      63920: gap of unknown length
*     63921      66860: contig of 2940 bp in length
*     66861      66960: gap of unknown length
*     66961      68743: contig of 1783 bp in length
*     68744      68843: gap of unknown length
*     68844      71603: contig of 2760 bp in length
*     71604      71703: gap of unknown length
*     71704      74491: contig of 2788 bp in length
*     74492      74591: gap of unknown length
*     74592      77206: contig of 2615 bp in length
*     77207      77306: gap of unknown length
*     77307      80580: contig of 3274 bp in length
*     80581      80680: gap of unknown length
*     80681      83678: contig of 2998 bp in length
*     83679      83778: gap of unknown length

```

```

*      83779      86973: contig of 3195 bp in length
*      86974      87073: gap of unknown length
*      87074      90192: contig of 3119 bp in length
*      90193      90292: gap of unknown length
*      90293      95916: contig of 5624 bp in length
*      95917      96016: gap of unknown length
*      96017      100345: contig of 4329 bp in length
*      100346      100445: gap of unknown length
*      100446      104770: contig of 4325 bp in length
*      104771      104870: gap of unknown length
*      104871      110000: contig of 5130 bp in length
*      110001      110100: gap of unknown length
*      110101      113116: contig of 3016 bp in length
*      113117      113216: gap of unknown length
*      113217      117352: contig of 4136 bp in length
*      117353      117452: gap of unknown length
*      117453      121355: contig of 3903 bp in length
*      121356      121455: gap of unknown length
*      121456      124627: contig of 3172 bp in length
*      124628      124727: gap of unknown length
*      124728      130091: contig of 5364 bp in length
*      130092      130191: gap of unknown length
*      130192      135015: contig of 4824 bp in length

Query Match
Best Local Similarity 100.0%, Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1268 TTGTGATGTCAGATTGTTA 1288
Db       143561 TTGTGATGTCAGATTGTTA 143541

RESULT 43
AC106294
LOCUS
DEFINITION
Rattus norvegicus clone CH230-76010, *** SEQUENCING IN PROGRESS
AC106294
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 191192)
Muzny,D.M., Adams,C., Adio-Odnola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimagne,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Boyle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunyah,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Hosmi,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

```

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunola, G.,  
Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Plums, E., Pu, L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,  
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,  
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y., F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 191192)  
Morley, K.C.

Direct Submission  
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 191192)  
Morley, K.C.

Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gl:18138815.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center Project name: GK0M  
Center Clone name: CH230-76010  
----- Summary Statistics  
Sequencing vector: Plasmid:  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap: version 0.990329  
Consensus quality: 125322 bases at least Q40  
Consensus quality: 132642 bases at least Q30  
Consensus quality: 138789 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 75 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1388: contig of 1388 bp in length  
1389  
1488: gap of unknown length  
1489  
2669: contig of 1181 bp in length  
2670  
2769: gap of unknown length  
2770  
3808: contig of 1039 bp in length  
3809  
3908: gap of unknown length  
3909  
4929: contig of 1021 bp in length  
4930  
5029: gap of unknown length  
5030  
6223: contig of 1194 bp in length  
6224  
6323: gap of unknown length  
6324  
7626: contig of 1303 bp in length  
7627  
7726: gap of unknown length  
7727  
8871: contig of 1145 bp in length  
8872  
8971: gap of unknown length  
8972  
10196: contig of 1225 bp in length  
10197  
10296: gap of unknown length  
10297  
11381: contig of 1085 bp in length  
11382  
11381  
11481: gap of unknown length

11482  
13166: contig of 1685 bp in length  
13167  
13266: gap of unknown length  
13267  
14374: contig of 1108 bp in length  
14375  
14474: gap of unknown length  
14475  
15707: contig of 1233 bp in length  
15708  
15807: gap of unknown length  
15808  
16862: contig of 1055 bp in length  
16863  
16962: gap of unknown length  
16963  
18060: contig of 1098 bp in length  
18061  
18160: gap of unknown length  
18161  
19511: contig of 1351 bp in length  
19512  
19612: gap of unknown length  
20795  
20795: contig of 1183 bp in length  
20796  
20894: gap of unknown length  
22725  
22725: contig of 1831 bp in length  
22726  
22825: gap of unknown length  
22826  
23832: contig of 1007 bp in length  
23833  
23932: gap of unknown length  
23933  
25112: contig of 1180 bp in length  
25113  
25212: gap of unknown length  
26435: contig of 1223 bp in length  
26436  
26535: gap of unknown length  
26536  
28046: contig of 1511 bp in length  
28047  
28146: gap of unknown length  
28147  
30041: contig of 1895 bp in length  
30042  
30141: gap of unknown length  
30142  
32406: contig of 2265 bp in length  
32407  
32506: gap of unknown length  
32507  
33667: contig of 1161 bp in length  
33668  
33767: gap of unknown length  
33768  
35177: contig of 1410 bp in length  
35178  
35277: gap of unknown length  
35278  
37020: contig of 1743 bp in length  
37021  
37120: gap of unknown length  
37121  
39350: contig of 2230 bp in length  
39351  
39450: gap of unknown length  
41003: contig of 1553 bp in length  
41004  
41103: gap of unknown length  
41104  
42456: contig of 1353 bp in length  
42457  
42556: gap of unknown length  
42557  
44708: contig of 2152 bp in length  
44709  
44808: gap of unknown length  
44809  
45863: contig of 1055 bp in length  
45864  
45963: gap of unknown length  
45964  
47550: contig of 1587 bp in length  
47551  
47650: gap of unknown length  
47651  
49436: contig of 1786 bp in length  
49437  
49536: gap of unknown length  
49537  
51357: contig of 1821 bp in length  
51358  
51457: gap of unknown length  
51458  
54021: contig of 2564 bp in length  
54022  
54121: gap of unknown length  
54122  
55866: contig of 2465 bp in length  
55867  
56686: gap of unknown length  
56687  
58137: contig of 1451 bp in length  
58138  
58237: gap of unknown length  
58239  
60536: contig of 2299 bp in length  
60537  
60636: gap of unknown length  
60637  
63842: contig of 3206 bp in length  
63843  
63942: gap of unknown length  
63943  
65950: contig of 2008 bp in length  
65951  
66050: gap of unknown length  
66051  
67766: contig of 1716 bp in length  
67767  
67866: gap of unknown length  
67867  
69616: contig of 1750 bp in length  
69617  
69716: gap of unknown length  
69717  
71766: contig of 2050 bp in length  
71767  
71866: gap of unknown length  
71867  
74342: contig of 2476 bp in length  
74343  
74442: gap of unknown length  
74443  
75742: contig of 1300 bp in length  
75743  
75842: gap of unknown length  
75843  
78294: contig of 2452 bp in length

```

* 78295 78394: gap of unknown length
* 78395 79952: contig of 1558 bp in length
* 79953 80052: gap of unknown length
* 80053 83228: contig of 3176 bp in length
* 83229 83328: gap of unknown length
* 83329 87055: contig of 3727 bp in length
* 87056 87155: gap of unknown length
* 87156 90022: contig of 2867 bp in length
* 90023 90122: gap of unknown length
* 90123 92263: contig of 2141 bp in length
* 92264 92363: gap of unknown length
* 92364 95642: contig of 3279 bp in length
* 95643 95742: gap of unknown length
* 95743 97352: contig of 1610 bp in length
* 97353 97452: gap of unknown length
* 97453 99828: contig of 2376 bp in length

Query Match 1.0%; Score 21; DB 2; Length 191192;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 938 TGTATGCTCTGCTCCAGG 958
DB 113832 TGTATGCTCTGCTCCAGG 113852

RESULT 44
AC094442/C
LOCUS AC094442 197174 bp DNA linear HTG 10-JUL-2002
DEFINITION Rattus norvegicus clone CH230-4F5, *** SEQUENCING IN PROGRESS ***
ACCESSION AC094442.3 GI:21716483
VERSION AC094442
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 197174)
Muzny,D.M., Adams,C., Adio-Oduola,B., Al-Osman,F.R., Allen,C.,
Albriocks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Blinze,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burreli,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,D., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kritovoy,C., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,M., Nickerson,E., Nwokenwo,S., Oguh,M., Okunolu,G.,
Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Pince,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rojle,M., Ruiz,S., Savery,G.,
Schierer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 197174)
Worley,K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 197174)
Worley,K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941170.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMP
Center clone name: CH230-4F5
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 158750 bases at least Q40
Consensus quality: 163771 bases at least Q30
Consensus quality: 168185 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1093: contig of 1093 bp in length
* 1094 1193: gap of unknown length
* 1194 2298: contig of 1105 bp in length
* 2299 2398: gap of unknown length
* 2399 4247: contig of 1849 bp in length
* 4248 4347: gap of unknown length
* 4348 5913: contig of 1566 bp in length
* 5914 6013: gap of unknown length
* 6014 7073: contig of 1060 bp in length
* 7074 7173: gap of unknown length
* 7174 8519: contig of 1346 bp in length
* 8520 8619: gap of unknown length
* 8620 9903: contig of 1284 bp in length
* 9904 10003: gap of unknown length
* 10004 11377: contig of 1374 bp in length
* 11378 11477: gap of unknown length
* 11478 12943: contig of 1466 bp in length
* 12944 13043: gap of unknown length
* 13044 14531: contig of 1488 bp in length
* 14532 14631: gap of unknown length
* 14632 16736: contig of 2105 bp in length
* 16737 16836: gap of unknown length
* 16837 18076: contig of 1240 bp in length
* 18077 18176: gap of unknown length
* 18177 19513: contig of 1337 bp in length

```

19514 19513: gap of unknown length  
 19614 20718: contig of 1105 bp in length  
 20719 20818: gap of unknown length  
 20819 22279: contig of 1461 bp in length  
 22280 22379: gap of unknown length  
 22380 24578: contig of 2199 bp in length  
 24579 24678: gap of unknown length  
 24679 26727: contig of 2049 bp in length  
 26728 26827: gap of unknown length  
 26828 28392: contig of 1465 bp in length  
 28393 29900: contig of 1508 bp in length  
 29901 30000: gap of unknown length  
 30001 31343: contig of 1343 bp in length  
 31344 33191: contig of 1748 bp in length  
 33192 33291: gap of unknown length  
 33292 35020: contig of 1729 bp in length  
 35021 35120: gap of unknown length  
 35121 36637: contig of 1517 bp in length  
 36638 36737: gap of unknown length  
 36739 38139: contig of 1402 bp in length  
 38140 38239: gap of unknown length  
 38240 40561: contig of 2322 bp in length  
 40562 40661: gap of unknown length  
 40662 42493: contig of 1832 bp in length  
 42494 42593: gap of unknown length  
 42594 44303: contig of 1710 bp in length  
 44304 44403: gap of unknown length  
 44404 46466: contig of 2063 bp in length  
 46467 46566: gap of unknown length  
 46567 48004: contig of 1438 bp in length  
 48005 48104: gap of unknown length  
 48105 50303: contig of 2199 bp in length  
 50304 50403: gap of unknown length  
 50404 53233: contig of 2830 bp in length  
 53234 53333: gap of unknown length  
 53334 55265: contig of 1932 bp in length  
 55266 55365: gap of unknown length  
 55366 57056: contig of 1691 bp in length  
 57057 57156: gap of unknown length  
 57157 59608: contig of 2452 bp in length  
 59609 59708: gap of unknown length  
 59709 63931: contig of 4223 bp in length  
 63932 64031: gap of unknown length  
 64032 66878: contig of 2847 bp in length  
 66879 66978: gap of unknown length  
 66979 69578: contig of 2600 bp in length  
 69579 71920: gap of unknown length  
 71921 72020: contig of 2242 bp in length  
 72021 74778: contig of 2758 bp in length  
 74779 77904: gap of unknown length  
 77905 78004: contig of 3026 bp in length  
 78005 81637: contig of 3633 bp in length  
 81638 81737: gap of unknown length  
 81738 83825: contig of 2088 bp in length  
 83826 83925: gap of unknown length  
 83926 86560: contig of 2635 bp in length  
 86561 86660: gap of unknown length  
 86661 88789: contig of 2129 bp in length  
 88790 88889: gap of unknown length  
 88890 90661: contig of 1772 bp in length  
 90662 90761: gap of unknown length  
 90762 94063: contig of 3302 bp in length  
 94064 94163: gap of unknown length  
 94164 97840: contig of 3677 bp in length  
 97841 97940: gap of unknown length  
 97941 100540: contig of 2600 bp in length  
 100541 103145: gap of unknown length  
 103146 103245: gap of unknown length

Query Match 1.0% Score 21: DB 2: Length 197174;  
 Best Local Similarity 100.0% Pred No. 5;  
 Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1220 TGCTTCACAGTGCAGAGCTCA 1240  
 Db 161075 TGCTTCACAGTGCAGAGCTCA 161055

RESULT 45  
 AC096705  
 LOCUS  
 DEFINITION  
 Rattus norvegicus clone CH230-218, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 70 unordered pieces.  
 AC096705  
 VERSION  
 KEYWORDS  
 HTG: HTGS\_PHASE1.  
 SOURCE  
 ORGANISM  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 212255)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barberia,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,  
 Homsi,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolyet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounseged,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Merlindade,A., Martinez,E.,  
 Massey,E., Mawhney,E., McLeod,M.P., Meadow,M., Mel,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okunou,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Saverly,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,  
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
 Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.



TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 212255)  
Morley, K. C.  
Submitted (23-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 212255)  
Morley, K. C.  
Direct Submission  
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 24, 2002 this sequence version replaced gi:15799569.

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information -----  
Center project name: TUSU  
Center clone name: CH230-218  
----- Summary Statistics -----  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 154690 bases at least Q40  
Consensus quality: 161468 bases at least Q30  
Consensus quality: 166127 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_drift\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 70 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1122: contig of 1122 bp in length  
1123  
1222: gap of unknown length  
1223  
2420: contig of 1198 bp in length  
2421  
2520: gap of unknown length  
2521  
3521: contig of 1001 bp in length  
3522  
3621: gap of unknown length  
3622  
4871: contig of 1250 bp in length  
4872  
4972  
6044: contig of 1073 bp in length  
6045  
6144: gap of unknown length  
6145  
7592: contig of 1448 bp in length  
7593  
7692: gap of unknown length  
7693  
8942: contig of 1250 bp in length  
9042: gap of unknown length  
9043  
10562: contig of 1520 bp in length  
10563  
10662: gap of unknown length  
10663  
11686: contig of 1024 bp in length  
11687  
11786: gap of unknown length  
11787  
13151: contig of 1365 bp in length  
13152  
13251: gap of unknown length  
13252  
14683: contig of 1432 bp in length  
14684  
14783: gap of unknown length  
14784  
15811: contig of 1028 bp in length  
15812  
15911: gap of unknown length  
15912  
17238: contig of 1327 bp in length  
17239  
17339: gap of unknown length  
17339  
18470: contig of 1132 bp in length  
18471  
18570: gap of unknown length  
18571  
19730: contig of 1160 bp in length  
19731  
19830: gap of unknown length  
20908: contig of 1078 bp in length  
20909  
21008: gap of unknown length

21009  
22441  
22541  
24411  
24510  
24510: gap of unknown length  
25900: contig of 1390 bp in length  
25901  
26000: gap of unknown length  
26001  
27471: contig of 1471 bp in length  
27472  
27571: gap of unknown length  
27572  
29135: contig of 1564 bp in length  
29136  
29236: gap of unknown length  
29237  
31590: contig of 2355 bp in length  
31591  
31690: gap of unknown length  
32914: contig of 1224 bp in length  
32915  
33014: gap of unknown length  
34913: contig of 1899 bp in length  
34914  
35013: gap of unknown length  
35014  
36934: contig of 1921 bp in length  
36935  
37034: gap of unknown length  
38412: contig of 1378 bp in length  
38413  
38512: gap of unknown length  
39754: contig of 1242 bp in length  
39755  
39854: gap of unknown length  
39855  
41115: contig of 1261 bp in length  
41216  
42490: contig of 1275 bp in length  
42491  
42591  
44367: contig of 1777 bp in length  
44368  
44467: gap of unknown length  
46114: contig of 1647 bp in length  
46214: gap of unknown length  
47502: contig of 1288 bp in length  
47602: gap of unknown length  
50193: contig of 2591 bp in length  
50293: gap of unknown length  
53097: contig of 2804 bp in length  
53197: gap of unknown length  
55497: contig of 2300 bp in length  
55597: gap of unknown length  
55971: contig of 1374 bp in length  
56972  
57071: gap of unknown length  
57072  
59081: contig of 2010 bp in length  
59181: gap of unknown length  
61650: contig of 2469 bp in length  
61651  
61750: gap of unknown length  
63734: contig of 1884 bp in length  
63735  
63735: gap of unknown length  
65282: contig of 1548 bp in length  
65382: gap of unknown length  
67312: contig of 1930 bp in length  
67412: gap of unknown length  
69737: contig of 2325 bp in length  
69837: gap of unknown length  
72845: contig of 3008 bp in length  
72845: gap of unknown length  
72945: gap of unknown length  
75984: contig of 3039 bp in length  
76084: gap of unknown length  
78508: contig of 2424 bp in length  
78608: gap of unknown length  
81574: contig of 2366 bp in length  
81575  
81675: gap of unknown length  
84191: contig of 2517 bp in length  
84291: gap of unknown length  
86431: contig of 2142 bp in length  
86434  
86533: gap of unknown length  
86534  
89649: contig of 3116 bp in length  
89650  
93795: contig of 4046 bp in length  
93795: gap of unknown length  
93796  
97729: contig of 3834 bp in length  
97829: gap of unknown length  
97830  
101606: contig of 3777 bp in length  
101706: gap of unknown length  
104900: contig of 3194 bp in length

Query Match	Best Local Similarity	1.0%; Score 21; DB 2; Length 212255;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 923 AGAAGCCAGAGCTGTATG 943		
DB 167067 AGAAGCCAGAGCTGTATG 167087		
RESULT 46		
AL592564	750 bp DNA	linear STS 16-JUL-2001
DEFINITION	Meleagris galllopavo DNA; STS TC0252, sequence tagged site.	
AL592564		
VERSION	AL592564.1 GI:14715682	
KEYWORDS	STS.	
SOURCE	turkey	
ORGANISM	Meleagris galllopavo	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Archosauaria; Aves; Neognathae; Galliformes; Meleagrididae;	
JOURNAL	Meleagris.	
REFERENCE	1 (bases 1 to 750)	
AUTHORS	Morrice,D.	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 750)	
AUTHORS	Morrice,D.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-JUL-2001) Morrice D., Roslin Institute, Roslin,	
COMMENT	Midlothian, EH25 9PS, UK david.morrice@bsrc.ac.uk	
	Vector pBLUESCRIPT SK+	
	R. Site 1 Ecor1	
	R. Site 2 Ecor1	
	Contact: David Morrice	
	Dept. Genomics and Bioinformatics	
	Roslin Institute, Roslin	
	Midlothian, EH25 9PS, UK	
	Tel: +44 (0) 131 527 4200	
	Fax: +44 (0) 131 440 0434	
	Email: david.morrice@bsrc.ac.uk	
	Seq primer 17.	
FEATURES	Location/Qualifiers	
source	1..750	
STS	/organism="Meleagris galllopavo"	
	/db_xref="taxon:9103"	
	1..750	
BASE COUNT	/standard_name="TC0252"	
ORIGIN	215 a 163 c 143 g 207 t 22 others	
Query Match	1.0%; Score 20; DB 11; Length 750;	
Best Local Similarity	100.0%; Pred. No. 18;	
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 275 TGAGTGCATTACACACAC 294		
DB 424 TGAGTGCATTACACACAC 443		
RESULT 47		
HSN800587	1065 bp mRNA	linear PRI 18-FEB-2000
LOCUS	Homo sapiens mRNA; cDNA DKFZp564J1762 (from clone DKFZp564J1762).	
DEFINITION	AL080086	
ACCESSION	AL080086.1 GI:5262502	
VERSION		
KEYWORDS		
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 1065)	

AUTHORS	Blum, H., Baumanns, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUN-1999) MIPS, Am Klopferstr. 18a, D-82152 Martinsried, GERMANY
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by LMU (Ludwig Maximilians University, Munich/Germany) within the cDNA sequencing consortium of the German Genome Project.
FEATURES	This clone (DKFZp564J1762) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Further information about the clone and the sequencing project is available at <a href="http://www.mips.biochem.mpg.de/proj/cDNA/">http://www.mips.biochem.mpg.de/proj/cDNA/</a> .
SOURCE	Location/Qualifiers
1. 1065	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="DKFZp564J1762"	
/tissue_type="brain"	
/clone_id="564 (synonym: hfbr2). Vector: pAMPI; host	
x1-blue; sites NotI + SalI"	
/dev_stage="fetal"	
BASE COUNT	989 268 c 311 g 211 t
ORIGIN	
polyA_site	
Query Match	1.0%; Score 20; DB 9; Length 1065;
Best Local Similarity	100.0%; Pred. No. 18;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1575 CACATGAGCAGTGTCCCG 1594
Db	121 CACATGAGCAGTGTCCCG 140
RESULT 48	
AF231711/c	
LOCUS	AF231711 1680 bp mRNA linear VRT 01-MAR-2000
DEFINITION	Xenopus laevis 7-transmembrane receptor frizzled-1 mRNA, complete cds.
ACCESSION	AF231711
VERSION	AF231711.1 GI:7110523
KEYWORDS	
SOURCE	
ORGANISM	Xenopus laevis.
	Xenopus laevis.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE	1 (bases 1 to 1680)
AUTHORS	Brown, J.D., Hallagan, S.E., McGrew, L.L., Miller, J.R. and Moon, R.T.
TITLE	The maternal Xenopus beta-catenin signaling pathway, activated by frizzled homologs, induces goosecoid in a cell non-autonomous manner
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1680)
AUTHORS	Brown, J.D., Hallagan, S.E. and Moon, R.T.
TITLE	Direct Submission
JOURNAL	Submitted (04-FEB-2000) Pharmacology and HHMI, University of Washington, Seattle, WA 98195, USA
FEATURES	Location/Qualifiers
SOURCE	1. 1680
/organism="Xenopus laevis"	
/db_xref="taxon:8335"	
1. 1680	
/codon_start=1	
/product="7-transmembrane receptor frizzled-1"	
/protein_id="AA36979.1"	
/db_xref="GI:7110524"	
/translation="MKSHLQRCSAQCTRGSSLLSLVCLSEGYNGEKGISIPDHGYCOPISLPCDIAVNOTIMPLDHTNEDAGLEHOFYPLVKOCSPELKFLCSITVAPCTVEQALPCRCSCDRARQCEALMNFQWPEBSLCEKFPINGAE	
LCGQNTTEGTPTPAVPEWTNLSKRIYVYDKRCPCPALVPAVNTVHFGKDCGAP	



gene  
complement(2964. .3692)  
/gene="g1db\_L"  
/note="g1db: The second copy of g1db gene is located in  
the right arm of Ladr"  
complement(2964. .3692)  
/gene="g1db\_L"  
/codon\_start=1  
/transl\_table=4  
/product="glucose inhibited division protein B"  
/protein\_id="AA36756.1"  
/db\_xref="GI:7109685"  
/translation="MLKIKDSLGKFNLTLDKQIEDIAFLEEIHVSNOLEFNTGYK  
TELLAEMLGQVILLASQSLVIFSNOSLIDGTGIGLIIKIIPOLYLVWD  
NAKKTITFNEVYIKLNFQTVGPAISREDNPFLKYGCPDVAQVSAVSLANLE  
GTOLKINGOIIHKSRYOEIEEFKAKHSLDLGLAFNNLIHYOPNSIFLVNFTYNNKKA  
ALAPKTFREMSKIKRELIDDKH"  
complement(3685. .5520)  
/gene="g1da\_L"  
/note="g1da: The second copy of g1da gene is located in  
the right arm of Ladr"  
complement(3685. .5520)  
/gene="g1da\_L"  
/codon\_start=1  
/transl\_table=4  
/product="glucose inhibited division protein A"  
/protein\_id="AA36757.1"  
/db\_xref="GI:7109686"  
/translation="MNFPIVVGAGHAGLEAFLISLKNKRYLCLDRKYVANCPCNP  
SVGSAKQIVIREIDALGIGIALADSYALQRIKILNSKSGPVOCLEFQIDKRYKKMK  
FLEQDNNENNELVEGEVTEVINKNDPATGVIMGVKKLEAGAVIITGVYKSLFSES  
CDVQNEPEEGEKNSNNLSEMPKVGAFELIRKGTGPRIKDSIDVSNQIQEGNCTE  
EIFYHMSKNKRTIDELPCYLIHTTEELHKIINDNLHSANSGNITGVCRTPSIE  
DKIVFPPKHQPILEPESLELGVYVIGFSTSDISVQDKIIRLGLPKAAVIRK  
GAIEYDAINPIQIPLPSLESLVKNLFFPAGQINSTGYEALAGGLAGIANOKIKR  
KEPILSDDEAYEIVGIMDIDVTKGTDPYRLTSAEYRELANDVYLDRIQCYEII  
GTSKKQDILVQNQLEKNKRLIEFLKDKVAYRLTAAHTNNTNLSYEFLKREIKI  
IELKLEFDVSNDELKLNIEITVEXEKGITKESRIVNLSLSIKIPDIIYDK  
VONSLIEDIKLNKKPILNLAQRISSINADIISLKLHLEQNA"  
5771.6259  
CDS  
/note="ORF162\_L: The second copy of ORF162 CDS is located  
in the right arm of Ladr"  
/codon\_start=1  
/transl\_table=4  
/product="unknown"  
/protein\_id="AA36758.1"  
/db\_xref="GI:7109687"  
/translation="MKLSLKSGLTREQKLEKISLTDKELFTYIIILFGYELFVVMNA  
FLIDVACNEFTYNPNAETSLTMAGWSQSFPEKNGDSITTAATNWTIFEFRCISFVAG  
WIGLGHKKKTVLTMGLMALSPFIYVAVYPPGGNNALVYLSNQIIFYQKQDLSGSS  
LT"  
6369.7349  
CDS  
/note="ORF326\_L: The second copy of ORF326 CDS is located  
in the right arm of Ladr"  
/codon\_start=1  
/transl\_table=4  
/product="unknown"  
/protein\_id="AA36759.1"  
/db\_xref="GI:7109688"  
/translation="MKRTYLLSVINSFGFNIGAAFSGLFTYSMTIKMEAVKSVYVIC  
AAFLIFATFPIVPIRKEKTEVLPKEXKTPKPEPTIMSTLKNKNAVIMKCLMAVAMLY  
AVVALLTCTYSNSVIGTSPRYVNGKLFPEVNGTSTKIAFLVGLVACLFYIAPPGKRTIRYR  
SRFLMTQYSLCMFEYVAVVGLFPGVANNICGLAIGEIIASFLAGIPLWGLQGVLLP  
HEFKASQSKQVQFGIIGWGYMLYXTMSDILSLVIESFPAKSYVDSINPASITVA  
FAYVYIISMVPLILVLPKSGRIYNGEWPVLTQDKMPMSYNFYKQDIFK"  
complement(7414. .8028)  
/gene="def\_L"  
/note="def: The second copy of def gene is located in the  
right arm of Ladr"  
complement(7414. .8028)  
/gene="def\_L"  
/EC\_number="3.5.1.31"  
/codon\_start=1  
/transl\_table=4

```

Query Match      1.0%; Score 20; DB 1; Length 17411;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1265 ATATTGTTGATGTCAGATT 1264
         |||||||||||||||||||
Db       1556 ATATTGTTGATGTCAGATT 1575

RESULT 50
AC116980
LOCUS     AC116980
DEFINITION Dictyostelium discoideum chromosome 2 map 1408084-1432089 strain
AA4, *** SEQUENCING IN PROGRESS ***, In ordered pieces.
AC116980
VERSION   AC116980.1 GI:20042935
KEYWORDS  HTG: HTGS PHASE2.
SOURCE    Dictyostelium discoideum.
ORGANISM  Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 23461)
AUTHORS   Gloeckner,G., Eichinger,L., Szafarski,K., Pachebat,J., Dear,P.,
          Lehmann,R., Baumgart,C., Parra,G., April,J.F., Gulgo,R., Kumpf,K.,
          Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
          Noegel,A.A.
          Sequence and Analysis of Chromosome 2 of Dictyostelium
          Unpublished
          The Dictyostelium Genome Sequencing Consortium
          2 (bases 1 to 23461)
          Baumgart,C.
          Direct Submission
          Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
          CBS predictions from Genaid may contain errors. Further information
          is available from IMB Jena, Department of Genome Analysis
          (http://genome.imb-jena.de/dictyostelium/)
          and the University Cologne, Institute for Biochemistry I
          (http://www.uni-koeln.de/dictyostelium/project.shtml)
          Funding
          Agency : Deutsche Forschungsgemeinschaft (DFG).
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          Location/Qualifiers
            1..23461
               /organism="Dictyostelium discoideum"
               /strain="AX4"
               /db_xref="taxon:44699"
               /chromosome="2"
               /map="1408084-1432089"
               complement(2324..2590)
               /note="ORF_ID:dd_01691"
               /codon_start=1
               /product="hypothetical protein"
               /protein_id="AA08754.1"
               /db_xref="GI:20042936"
               /translation="MEPCVILENPLSEPRFLCTNDLVEYNHQSSEGHYILETF
               KGRFKNSHLITITFENNGGGGSKYKNITLTITNLRIVYKOTKE"
               join(2792..2868,2908..3049)
               /note="ORF_ID:dd_01692"
               /codon_start=1
               /product="hypothetical protein"
               /protein_id="AA08755.1"
               /db_xref="GI:20042937"
               /translation="MNAICIDAYFNMFKILITQYTTIDELLILSLISLFEFVYVDD
               DYUUDSGSNHNHNHNHNHNHNHNHDI"
               complement(join(3380..5292,5378..5641,5910..5937))
               /note="ORF_ID:dd_01693"
               /codon_start=1
               /product="hypothetical protein"

```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 09:08:54 ; Search time 338 seconds  
(without alignments)  
9462.933 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgtcactactatg.....ttgaagtgtcactactag 2049

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2-6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2-6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*
- 3: /cgn2-6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2-6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2-6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2-6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2-6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2-6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2-6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2-6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2-6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2-6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2-6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2-6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.8	21.1	1194	10	US-09-739-907-49 Sequence 49, Appl
2	430.2	21.0	1222	10	US-09-739-907-31 Sequence 31, Appl
3	121.2	5.9	1941	10	US-09-777-710A-2 Sequence 2, Appl
4	121.2	5.9	2013	10	US-09-777-710A-16 Sequence 16, Appl
5	72.2	3.5	272	10	US-09-878-574-9145 Sequence 9145, Ap
6	39.8	1.9	713	10	US-09-910-943-91 Sequence 91, Appl
7	37.4	1.8	462	10	US-09-954-456-1057 Sequence 1057, Ap
8	36	1.7	170	10	US-09-878-574-8039 Sequence 8039, Ap
9	34	1.7	14155	9	US-10-108-605-102 Sequence 107, App
10	33.6	1.6	1898	10	US-09-822-849A-310 Sequence 310, App
11	33.4	1.6	415	9	US-10-186-846-2496 Sequence 2496, Ap
12	33.2	1.6	1137	9	US-10-184-644-518 Sequence 518, App
13	33.2	1.6	1137	9	US-10-184-634-518 Sequence 518, App
14	33	1.6	1523	9	US-10-184-644-290 Sequence 290, App
15	33	1.6	1523	9	US-10-184-634-290 Sequence 290, App
16	33	1.6	7040	9	US-10-172-086-13 Sequence 13, Appl
17	32.8	1.6	912	9	US-09-738-626-924 Sequence 924, App
18	32.8	1.6	1473	9	US-09-738-626-923 Sequence 923, App
19	32.8	1.6	2643	10	US-09-815-242-7116 Sequence 7116, Ap

20	32.8	1.6	4929	9	US-09-738-626-3394 Sequence 3394, Ap
21	32.8	1.6	8354	9	US-10-125-818-1 Sequence 1, Appl
22	32.8	1.6	1503841	9	US-09-946-807-1 Sequence 1, Appl
23	32.8	1.6	1503841	10	US-09-795-668-1 Sequence 1, Appl
24	32.8	1.6	1503841	10	US-09-795-668-1 Sequence 1, Appl
25	32.8	1.6	1830121	9	US-10-329-960-1 Sequence 1, Appl
26	32.8	1.6	3309400	9	US-09-738-626-1 Sequence 1, Appl
27	32.8	1.6	3309400	9	US-09-738-626-1 Sequence 1, Appl
28	32.6	1.6	471	9	US-10-066-543-2224 Sequence 2224, Ap
29	32.6	1.6	1149	12	US-10-044-090-9 Sequence 9, Appl
30	32.6	1.6	1242	9	US-10-083-357-601 Sequence 601, App
31	32.6	1.6	2393	10	US-09-964-824A-281 Sequence 281, App
32	32.6	1.6	3316	9	US-10-242-056-58 Sequence 58, Appl
33	32.6	1.6	3752	9	US-10-098-841-55 Sequence 55, Appl
34	32.6	1.6	3752	9	US-10-037-270-691 Sequence 691, App
35	32.4	1.6	7287	9	US-09-938-842A-1019 Sequence 1019, Ap
36	32.4	1.6	20300	9	US-09-939-202A-3 Sequence 3, Appl
37	32.4	1.6	20556	10	US-09-880-107-3945 Sequence 3945, Ap
38	32.2	1.6	547	9	US-09-918-995-27398 Sequence 27398, A
39	32.2	1.6	1570	9	US-09-938-842A-4436 Sequence 4436, Ap
40	32.2	1.6	19553	9	US-10-092-154-1425 Sequence 1425, Ap
41	32.2	1.6	19553	10	US-09-764-847-1425 Sequence 1425, Ap
42	32.2	1.6	70768	9	US-10-135-322-13 Sequence 13, Appl
43	32	1.6	238	10	US-09-878-574-10358 Sequence 10358, A
44	32	1.6	277	10	US-09-878-574-14108 Sequence 14108, A
45	32	1.6	383	10	US-09-878-574-1049 Sequence 1049, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-739-907-49  
Sequence 49, Application US/09739907  
Patent No. US20010012889A1

GENERAL INFORMATION:  
APPLICANT: Rosen et al.

TITLE OF INVENTION: 36 Human Secreted Proteins  
FILE REFERENCE: P2022p1

CURRENT APPLICATION NUMBER: US/09/739, 907  
CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: 09/348, 457  
PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: 60/070, 567  
PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070, 692  
PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070, 704  
PRIOR FILING DATE: 1998-01-07

PRIOR APPLICATION NUMBER: 60/070, 658  
PRIOR FILING DATE: 1998-01-07

NUMBER OF SEQ ID NOS: 196  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 49

LENGTH: 1194  
TYPE: DNA  
ORGANISM: Homo sapiens

US-09-739-907-49

Query Match  
Best Local Similarity 81.1%; Score 431.8; DB 10; Length 1194;  
Matches 502; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

1 TTCAGTGCCTACTACATTGGCGTGGAGAGCGGATTGCACAGCAAGTGCAGATATC 60  
|||||  
426 TTCAGTGCCTACCAATGCTGTGGAGAGTCTGCACAGCAAGTGCAGATATC 485  
|||||  
61 AATTGATGGGCTATGCGCAAAATGCGCAGATGACGCGGCTCCACAGGCTGTATC 120  
|||||  
486 AATTGATGGGCTATGCGCAAAATGCGCAGATGACGCGGCTCCACAGGCTGTATC 545  
|||||  
121 AGCCGTGCTTTATCTTGGCGGATCCAGATGGGCTCAATGCAATGCAATTTGTGAGCGTG 180  
|||||

Db 546 AGTCGTGCTTCAATCATGCGAAGAACTGATGGCTCCATGCAACAGTGTGTCAGCATC 605  
Qy 181 GAACATGATGATGATTTCCCAACGACTGAGGTGGAGGTCCCTGAAGAGACTAGAGATAA 240  
Db 606 GACATGGCATGGTATCCCAAGGCTCAGGCTGGAGGTCTCTGAACACACTGCAGAGATAA 665  
Qy 241 TATGGCTCTGTATGACAGACAAATGTTATCTGTAGTCCATTCACACACTCTGGC 300  
Db 666 TATGGCTCTGTATGACAGACAAATGTTATCTGTAGTCCATTCACACTCTGTAGT 725  
Qy 301 CCAGCAGGGTTTTCCAAATATACACTATATCTGCGCAGCGAGGATTCAGACACCGG 360  
Db 726 CCGCAGAGATTTTCCAGTATACCGGTGTTGTAATGCGAGTGAGGATTTAGCAATCAA 785  
Qy 361 ACCTTCAGTACATAGTCTCTGGGATCATGAGAGCATGATATGCTCAGACAAATCTT 420  
Db 786 ACTTTCAGCAGATTTTCCAGTATACCGGTGTTGTAATGCGAGTGAGGATTTAGCAATCAA 845  
Qy 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGACATCAACCGAAGC 480  
Db 846 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGACATCAACCGAAGT 905  
Qy 481 CCCTCTCTTACCTTCTGTAATCCACAGTACAGAGACAGCAAGTATCTTCAACACAGAC 540  
Db 906 CCGTATTTCTTACCTTCAAAATCCGCACTCAGAGAGCAAGTATTTCTTCAATACAGAC 965  
Qy 541 AAGCAAAATGCTGCTTGAACAGTGGATTGTAATGAGAGACAGTGGGCTTATCAGC 600  
Db 966 AAGCAAAATGCTGCTTGAACAGTGGATTGTAATGAGAGATGACTTGGGCTTATCAGT 1025  
Qy 601 TGGTTGCCATCCACCCCG 619  
Db 1026 TTTTCATTTGCAAGTCTG 1044

RESULT 2  
US-09-739-907-31  
; Sequence 31, Application US/09739907  
; Patent No. US20010012889A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 36 Human Secreted Proteins  
; FILE REFERENCE: P2022P1  
; CURRENT APPLICATION NUMBER: US/09/739,907  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 09/348,457  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: 60/070,567  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: 60/070,692  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: 60/070,704  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: 60/070,658  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 196  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 31  
; LENGTH: 1222  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-739-907-31

Query Match 21.0%; Score 430.2; DB 10; Length 1222;  
Best Local Similarity 80.9%; Pred. No. 8.6e-132;  
Matches 501; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 1 TTAGAGGCTACTACATGCGCTTGGGAGACGCGATTCACAGACAGTGTCAATATC 60  
Db 449 TTACAGGCTACTACATGCGCTTGGGAGACGCGATTCACAGACAGTGTCAATATC 508  
Qy 61 AATTGATGGCTATGGCAAAATGGCCAAATGACAGGGGTCTCCCTACACAGGCTTTC 120  
|||||

Db 509 AATTGATGGCTATGGCAAAATCCGCGAGAAATGACACAGGCGATCTCCACAGGCTATAC 568  
Qy 121 AGCGGTGCTTTTATCTTGGGAGATCCAGATGGGTCAATATGCAATGCAATTTGTAGCGTG 180  
Db 569 AGTGTGCTTTCATTCATGGGAGAAACCTGATGGGTCCATATGAAACAGTGTGTCAGATC 628  
Qy 181 GAACATGATGATTTCCCAACGACTGAGGTGGAGGTCTCTGAAGAGACTAGAGATAA 240  
Db 629 GACATGAGCATGGTATCCCAAGGCTCAGGCTGGAGGTCTCTGAACACACTGCAGAGATAA 688  
Qy 241 TATGGCTCTGTATGACAGACAAATGTTATCTGTAGTCCATTCACACACTCTGGC 300  
Db 689 TATGGCTCTGTATGACAGACAAATGTTATCTGTAGTCCATTCACACTCTGTAGT 748  
Qy 301 CCAGCAGGGTTTTCCAAATATACACTATATCTGCGCAGCGAGGATTCAGACACCGG 360  
Db 749 CCGCAGAGATTTTCCAGTATACCGGTGTTGTAATGCGAGTGAGGATTTAGCAATCAA 808  
Qy 361 ACCTTCAGTACATAGTCTCTGGGATCATGAGAGCATGATATGCTCAGACAAATCTT 420  
Db 809 ACTTTCAGCAGATTTTCCAGTATACCGGTGTTGTAATGCGAGTGAGGATTTAGCAATCAA 868  
Qy 421 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGACATCAACCGAAGC 480  
Db 869 AAACGAGCAAAATCTTTATCAACAAAGAAATGTTGCTAATGTGACATCAACCGAAGT 928  
Qy 481 CCCTCTCTTACCTTCTGTAATCCACAGTACAGAGACAGCAAGTATCTTCAACACAGAC 540  
Db 929 CCGTATTTCTTACCTTCAAAATCCGCACTCAGAGAGCAAGTATTTCTTCAATACAGAC 988  
Qy 541 AAGCAAAATGCTGCTTGAACAGTGGATTGTAATGAGAGACAGTGGGCTTATCAGC 600  
Db 989 AAGCAAAATGCTGCTTGAACAGTGGATTGTAATGAGAGATGACTTGGGCTTATCAGT 1048  
Qy 601 TGGTTGCCATCCACCCCG 619  
Db 1049 TTTTCATTTGCAAGTCTG 1067

RESULT 3  
US-09-777-710A-2  
; Sequence 2, Application US/09777710A  
; Patent No. US20020058305A1  
; GENERAL INFORMATION:  
; APPLICANT: OKINO, No. US20020058305A1omu et al.  
; TITLE OF INVENTION: CERAMIDASE GENE  
; FILE REFERENCE: 1422-0458P  
; CURRENT APPLICATION NUMBER: US/09/777,710A  
; CURRENT FILING DATE: 2001-02-07  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1941  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-777-710A-2

Query Match 5.9%; Score 121.2; DB 10; Length 1941;  
Best Local Similarity 49.7%; Pred. No. 8.6e-29;  
Matches 340; Conservative 0; Mismatches 338; Indels 6; Gaps 1;

Qy 9 CTACTACATTTGGCTTGGGAGAGCGGATTCACAGACAGAAATGTCAGATATCAATTTGAT 68  
Db 12 CTACCGCTTGGGCTTGGGAGAGCGGATTCACAGACAGAAATGTCAGATATCAATTTGAT 71  
Qy 69 GGGCTATGGCAAAATGCGCAAGATGACGCGGCTCTCACCAGCGGTGTTACGCGGTGC 128  
Db 72 GGGCTATGGCAAAATGCGCAAGATGACGCGGCTCTCACCAGCGGTGTTACGCGGTGC 131  
Qy 129 TTTTATCTTGGGAGATCCAGATGGGTCAAAATGCAATGCAATTTGTAGCGGTGAACTATG 188  
Db 132 CTTCGTATGAGAGAACGGCGACGCGGACGCTGCTAGCTCAACACGACGCTGGG 191  
|||||



OY		189	TATGATTGCCAACGACAGTGGGTGGAGTGCTCGTAAGAAGACTAGAAAGTAAATATGGCTC	248
Dd		192	GATGACCTTCACAGGCGCGTGACACCTGAAGGTTCTGTGGCCCGGCTCAACGGCAAGTACC	251
OY		249	TCTGATATGAAGAGACAATGTATCTGTGATGCTGCATTCACACACAATCGGGCCACAGG	308
Dd		252	TGTCTACGACAGGAACCAAGTATATGCTGGCCGACACCCACACCCAATTCGGGTCCGGGCG	311
OY		309	GTTTTTCCAAATATACACTCTATATATCTGCCCGCAGCGAGGATTCAGTACCGGACCTTTCA	368
Dd		312	CTTCCGCCACATACGGGATGTACAACTGTGCGTGTGCTTCCAGAAAAGACCTTCAA	371
OY		369	GTACATAGTCTCTGTGGATCATGAAGACATTGATAATAGCTCACACAAATCTTAAACCAGG	428
Dd		372	CGCCATCTCGACGGGCACTGTGCGCTCCATCAGCGGGGCCGAGGCTTAGGTTCAGCCCCG	431
OY		429	CMAATCTTTATCAACAAAGGAAATGTTGCTAATGTGCAATCAATCCGAAGCCCTTC	488
Dd		432	CCGCGCTGTTTCAACGGCAGCGGGAAGCTGGCAACGCCACCGCAACCTGTTGCTGTCTC	491
OY		489	TTACCTTGATCCACAGTCCAGAGAGCAAGATTTCTTAAACACAGACAAAGGAAT	548
Dd		492	GCACCTGAAGAATC-----CGGACATGCGCCGGCTTACGAGATGGCATGACCGCGAGAT	545
OY		549	GCTGGTCTTGAACATGCTGATTTGAATGSAAGAAGCTTGGGTCTTATCAGCTGCTTGC	608
Dd		546	GAGGTGTCTACGCTTCTGTGACGCCAACGGCAGAGCTGGCCGGCGATGATGTTGGTCCC	605
OY		609	CATCCACCCCGTAGCATGAACAAATAGCAACCACTTTGTAATATGACAATATGGCGTA	668
Dd		606	GGTGACACGACACTGTATGACCACAGCCCAATCATCCTGATCTCCCGACACAAGAGGCTA	665
OY		669	TGCGGCTTACCTTTTGAACAAGA	692
Dd		666	CGCCTCTTATCTAGTGGAGACAGA	689

```

RESULT 4
US-09-777-710A-16
: Sequence 16, Application US/09777710A
: Patent No. US20020058305A1
: GENERAL INFORMATION:
: APPLICANT: OKINO, No. US20020058305A1omw et al.
: TITLE OF INVENTION: CERAMIDASE GENE
: FILE REFERENCE: 1422-0458P
: CURRENT APPLICATION NUMBER: US/09/777,710A
: CURRENT FILING DATE: 2001-02-07
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 16
: LENGTH: 2013
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-16

```

Query Match	5.9%	Score 121.2	DB 10	Length 2013
Best Local Similarity	49.7%	Pred.No. 8.8e-29		
Matches 340	Conservative	0	Mismatches 338	Indels 6
				Gaps 1

QY 9 TTACTACATTTGGCGTTGGGAGAGCGGATTTGCACAGACAAAGTGTACAGATATCAATATTTGAT 68  
 Db 84 CTACCGCTTGGCCCTGGGCGAGCGGACATCCACCGGCAGCCGCCGAAGTGGCATGAT 143  
 QY 69 GGGCTATGGCAAAAATGGCCAGAAATGCACGGGGTCTCTCACAGGCTTTGACGGCTGC 128  
 Db 144 GGGTACTCTCTCCGAAACAGAAACCGCGCGCATCCACATGCGCCAGTTGGGCGGCTGC 203  
 QY 129 TTTTATCTTGGCGGATTCAGATGGGTCAATTCGAAATGGCATTTTGTAGGGTGGAACTATG 188  
 Db 204 CTTGTGATCGAGGAAGCGCCAGCGGACGTGCCCTGGGTCTACGTAAACACCGACCTGGG 263  
 QY 189 TATGATTTCCCAACGACTGAGGTTGAGGTCTCTGAAGAGACTAGAGATAAATATATGCTC 248

Db 264 GATGACCTTCCAGAGCCCGTCACCTGAAGTCTCTGGCCCGGCTCAAGCCGAAGTACCCCG 323

Qy 249 TCTGATCGAAGAGACAATGTTATCTCGTAGTGCCATTTCACACACACTCTGGCCCAAG 308

Db 324 TGCTACACACGAGAACCAACGTATGCTCCGCCACACCACACCCACTCTCCGGTCCGGGCG 383

Qy 309 GTTTTCCAAATATACACTATATATCTCGCCAGCGAGGATTTACAGCAACCGGACTTTTA 368

Db 384 CTTCTCCCACTACGGGATGTACAACTCTGTCGGTGTCTCCAGGAAAAAGACTTTCAA 443

Qy 369 GTACATAGTCTCTGGGATCTAAGAGACATTTGATATAGCTCACACAAATCTTAAACAG 428

Db 444 CGCCATCTCGACGGGCATCTGCTCCGCTCCATCGAGCGGGCCCAAGCCAGGTGTGAGGCCCG 503

Qy 429 CAAATCTTTATCAACAAAGAAATGTCTCTATGTGCAGATCAACCGAAGCCCTCTCTC 488

Db 504 CCGCTGTCTTAAGGACAGGGGAGCTGCGCAACGCAACCGCAACCTTGTGCTGTCTC 563

Qy 489 TTACCTCTGATTCACACATCCACAGTCAGAGAGACAAAGTATTTCTTAAACACAGACAAGAAAT 548

Db 564 GCACCTGAAGATC-----CGGACATTCGCGGCTATCGAGAGATGGCATGACCCCGAGAT 617

Qy 549 GCTGCTCTTGAACAGCTGTGAGATTTGAATGGAAAGACCTTGGGCTTATATACAGCTGTTTC 608

Db 618 GAGCGCTGTCACTCTGTGACGCCAACGGGACGCTGCGCGGCGCGATAGTGGTGTCC 677

Qy 609 CATCCACCCCGTAGACATGATCAATAGCAACCACTTTGAAATAGTACAAATATGGGCTA 668

Db 678 GGTGACACGACCTGTGATGATGACCAAGCCCAATCAGTGAATCTCCCGGACACAAGAGGCTA 737

Qy 669 TGGCGCTTACCTTTTGAAGACA 692

Db 738 GCGCTCTATCACTGGGAGACGA 761

```

RESULT 5
US-09-878-574-9145
; Sequence 9145, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 9145
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701102138H1
US-09-878-574-9145

```

[illegible]



```

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 31133
; CURRENT APPLICATION NUMBER: US/10/108,605
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO: 102
; LENGTH: 14155
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-108-605-102

```

```

Query Match 1.7% Score 34; DB 9; Length 14155;
Best Local Similarity 51.3%; Pred. No. 40;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```

```

QY 936 GCTGTATGCTCTCTCCCTCCAGGAGTGACCGGCCCACTGCTTGAGCTCACCAGTGGT 995
DB 10484 GCAGTTGGGCCAGGAATCGAAGACTTGAAGGCCCAAGTGGAGGCTGCTCCCACTTGGC 10543
QY 996 GAACATGACAGATGTGAGCGCTCCAGCTCAATGCCACACACAGTGAAGAGCTGAACC 1055
DB 10544 CAACAGCATCAAGTGGCGGCGCATTTTCAAGCCAAAGCAGTCTCTGAGACTGAAGACACC 10603
QY 1056 TGCCCTGGGCTACAGTTTGGCCGACGACCAATT 1089
DB 10604 GGAAGAAGCCAAAGTTGCTAGCCACTCGACCAAT 10637

```

```

RESULT 10
US-09-822-849A-310/C
; Sequence 310, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Mong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakhar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 310
; LENGTH: 1898
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-310

```

```

Query Match 1.6% Score 33.6; DB 10; Length 1898;
Best Local Similarity 57.7%; Pred. No. 12;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 1792 GTGAGAAATACGAGACTCTGTAGCTGACGAGATTAATGATATACAGATGCTCTGG 1851
DB 1705 GGGATGAAGAGATGACCTGTGGTGGCGGTGACAGCAGCAGTGTGAGTCTGATCTCAG 1646
QY 1852 GAGACGAGTTTATTTGGCACAAGAAATGACTGGCTGACCAA 1895
DB 1645 GGGACAGCTGACACAGCAGCTGGGAATCTGGGACGAGCACA 1602

```

```

RESULT 11
US-10-198-846-2496
; Sequence 2496, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2496
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 5
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-2496

```

```

Query Match 1.6% Score 33.4; DB 9; Length 415;
Best Local Similarity 65.3%; Pred. No. 4.6;
Matches 49; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```

```

QY 709 CTGCTGACAGAGGACGCTTGTAGCAGCTTGTGCTTCAATCTCGAGAGCTGTCA 768
DB 58 CCGCCCGGAGAGTCAATCTGAGAGCTGCGCTTCATCATCATCTCTCAAGACTCATCG 117
QY 769 CCCAATCTCTTGGC 783
DB 118 CCAAGCATCTCTGCG 132

```

```

RESULT 12
US-10-184-644-518
; Sequence 518, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guiney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO: 518
; LENGTH: 1137
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-518

```

[illegible]

```

OY      1205 AACCCAAACCAATCCTGCTTACAGTAGAGAGCTGCAGATACCACATCCTTGCGAACACAG 1266
           |||
DB       472 EQLCDFEEOCKAGEDEQACGTTDFESPREGCEDMSVGLQMRVRSADSOSSSNAACH 531
OY      1265 ATATTGTTCAGATTCAGATTGTTTACCGTTGGGTCCTTGCGCATTAAGCTGTATCCCTGGG 1322
           :|||:|
DB       532 FLSTGRAMGOLGAERARVLTPLLGPSGPCELHLAYLQSOPREVSINFEDRTCSVPGLH 591
OY      1335 AATTATACCAACCATGTCGGGAGAGACATTTTCGTAGAGCCAATTAAAAACAATTGGACATT 1388
           :|:::|:|
DB       592 SDTHRWVESKRPPDHTTGGGHFLPDTPTPLAGHSANHLSPQVPAAPTECLSFWH 651
OY      1385 ATGGGATGAGGATATATGACCCGTTGTTATCGCAGGCTGTACAGCAATPTTATACAACTACA 1444
           :|:|
DB       652 LHGPDIOTLRILAMRREGETHLMSSGTGNRMHEAMNTLSIQPSHAOYLLEGIARDG 711
OY      1445 TTACACATATGAGAATAACAGAGCTCACGCCGTACAGAGCAGACATCTCAATCTATGAC 1500
DB       712 YHGFWALDDVAVRPQCWAPNVCSPFEDSDCGSPSGOGILMRQANASGHIAAWGPPTDHT 771
OY      1505 CACACACCCCTGTCTGCATATACCAACACTCTTCAGAGACCTTGCTAAGCAATGCTACAG 1566
           :|:::|:|
DB       772 ETAGCHYVVVDITSPDALPRGQTASTJTSKEHRPLAOPACLTFWYHSLSPGLRVLYEBR 831
OY      1565 ACACAGTAGCCAACTGACAG 1586
           :|:|:|
DB       832 GRHOYLSLSAHGLAWRLGSMD 853

RESULT 14
US-10-184-644-290
; Sequence 290. Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184_644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 290
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-290

Query Match          1.6%, Score 33; DB 9; Length 1523;
Best Local Similarity 9.1%; Pred. No. 16;
Matches 57; Conservative 163; Mismatches 406; Indels 1; Gaps 1;

OY      465 GCAGATACAGCAAGCCCCCTCCTCTGTAACCTGTAATCCACAGTCAGAGAGCAAGGTA 524
           :|:|:|:|
DB       353 GLKSTSLVLGNKKITEPIAKGLFDGLVSLDLLLNANKINCLIRNVTFODLONLNLISLDY 412
OY      525 TTTCTTCAACACAGACAGAAATGCTGCTTGAACCTGATTTGAATGAGAGA 584
           :|:|:|:|
DB       413 NKLGTFISKGLFAPILOSIDTLHLADNPFCDCDLKMLADYLQDNPIETGACSSPRLIAN 472
OY      585 CTGGGCTTATCAAGCTGGTTTGGCCATCCACCCCGTGAGCATGAACAATGACCAACTT 644
           :|:|:|:|
```

Db 473 KRISOIKSKKRCSSGSEDEYRSRFSSECFMDLVCEKRCBEGTIVDC:SNOKIVRIPSHLPE 532  
OY 645 TGTGAATGATGACATATGCGCTTACGCTTTTGAGCAGAGAAGAAAGCAAGG 704  
Db 533 YVTDRLNDNEVSYLEATGIFKKLPNLK -INLSNNKIKEYREGAFDGAASVOELMTGN 591  
OY 705 CTATCGCTGACAGAGGACCGTTGTAGAGGCTTTGCTTCAATCAATCTCGAGAGCT 764  
Db 592 QLEIVHGVFRGLSGKTLMLRSLNLSVNDTFAGLSVRLSLYDRITITTPGATTT 651  
OY 765 GTCACCCCAACATTTCTGGCCCCGACATTTGTCAACACAGGAGGCTTTGTGACAACGACAA 824  
Db 652 LVSLSTINLSNPNFNCNHLMLGKWLKRRRIVSGNPRCQKPFELKEIPIDVVAIODEFTC 711  
OY 825 GAGACCTGTCCCAACGCTGGGCGCTAGCATGTGCATGGCCAGCGACCTGACAAAGACAT 884  
Db 712 DNESSCQLSPRCBQCTCMEYVRCNNKGLRALPRMPKDVITLIEGHNLAIVPREL 771  
OY 885 GTTGAAGACACACATTTATAGACGATCATCTATCAGAAGGCAAGAGAGCTGTATGC 944  
Db 772 SALRHLFLIDLNNISIMLTWTFNSMHSHTLILSYRLRCIPVHAENGRLSLRVLTFLH 831  
OY 945 CTCGCTCCAGAGAGGTGACCGGCCAGTCTTGACCTGACACGNGTGGTGAACATGAC 1004  
Db 832 GNDISSVEGSEFNDLTSLSHLALGTNPJLHCDCLRLMSEWYKAGYKEPFIARCSPEBMA 891  
OY 1005 AGATGTGACGCTGACGATCATGACACACACAGTGAAGCGTGAACCTGGCCCTGG 1064  
Db 892 DRLLITPTTHRQCKGPVDINIVAKNCNGLSSPCKNNGTCTQDPVELYRCACPYSYKGD 951  
OY 1065 CTACAGTTTTCGCCGAGGACACATTTGA 1091  
Db 952 CTVPINICIONPCQHGGTCHLSDSHKD 978

RESULT 15  
US-10-184-634-290  
: Sequence 290, Application US/10184634  
: Publication No. US20030068684A1  
: GENERAL INFORMATION:  
: APPLICANT: Baker, Kevin P.  
: APPLICANT: Chen, Jlan  
: APPLICANT: Desnoyers, Luc  
: APPLICANT: Goddard, Audrey  
: APPLICANT: Godowski, Paul J.  
: APPLICANT: Gurney, Austin L.  
: APPLICANT: Pan, James  
: APPLICANT: Smith, Victoria  
: APPLICANT: Watanabe, Colin K.  
: APPLICANT: Wood, William I.  
: APPLICANT: Zhang, Zemin  
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
: FILE OF INVENTION: ACIDS ENCODING THE SAME  
: FILE REFERENCE: P3430RIC217  
: CURRENT APPLICATION NUMBER: US/10/184,634  
: CURRENT FILING DATE: 2002-06-28  
: Prior Application removed - See file wrapper or Palm  
: NUMBER OF SEQ ID NOS: 612  
: SEQ ID NO 290  
: LENGTH: 1523  
: TYPE: PRT  
: ORGANISM: Homo Sapien  
US-10-184-634-290

Query Match 1.68, Score 33, DB 9, Length 1523;  
Best Local Similarity 9.1%; Pred. No. 16;  
Matches 57, Conservative 163, Mismatches 406, Indels 1, Gaps 1;

OY 465 GCAGATGACCGAGAGCCCTCTTACCTCTGATGACAGATGACAGAGAGAGAGAGAGTA 524  
Db 353 GLKSLTSLVLYGNKITEIAKLPDGLVSLQLLLLNANKINCLRVNTFODIONLNLISLYD 412  
OY 525 TTCTTCAACACAGACAGAAATGCTGCTTGAACGTGATTTGAATGAGAGAGA 584

Db 413 NKLOPISKGLPAPLQSIQTLHLAQNPFVCDCHLKLADYLDNPLETSGARCSSPRLAN 472  
OY 565 CTGGGCTTTATGACCTGTTTGCATCATCCACCCGCTGACATGAACATATACACACACTT 644  
Db 473 KRISOIKSKKRCSSGSEDEYRSRFSSECFMDLVCEKRCBEGTIVDC:SNOKIVRIPSHLPE 532  
OY 705 CTATCGCTGACAGAGGACCGTTGTAGAGGCTTTGCTTCAATCAATCTCGAGAGCT 764  
Db 592 QLEIVHGVFRGLSGKTLMLRSLNLSVNDTFAGLSVRLSLYDRITITTPGATTT 651  
OY 765 GTCACCCCAACATTTCTGGCCCCGACATTTGTCAACACAGGAGGCTTTGTGACAACGACAA 824  
Db 652 LVSLSTINLSNPNFNCNHLMLGKWLKRRRIVSGNPRCQKPFELKEIPIDVVAIODEFTC 711  
OY 825 GAGACCTGTCCCAACGCTGGGCGCTAGCATGTGCATGGCCAGCGACCTGACAAAGACAT 884  
Db 712 DNESSCQLSPRCBQCTCMEYVRCNNKGLRALPRMPKDVITLIEGHNLAIVPREL 771  
OY 885 GTTGAAGACACACATTTATAGACGATCATCTATCAGAAGGCAAGAGAGCTGTATGC 944  
Db 772 SALRHLFLIDLNNISIMLTWTFNSMHSHTLILSYRLRCIPVHAENGRLSLRVLTFLH 831  
OY 945 CTCGCTCCAGAGAGGTGACCGGCCAGTCTTGACCTGACACGNGTGGTGAACATGAC 1004  
Db 832 GNDISSVEGSEFNDLTSLSHLALGTNPJLHCDCLRLMSEWYKAGYKEPFIARCSPEBMA 891  
OY 1005 AGATGTGACGCTGACGATCATGACACACACAGTGAAGCGTGAACCTGGCCCTGG 1064  
Db 892 DRLLITPTTHRQCKGPVDINIVAKNCNGLSSPCKNNGTCTQDPVELYRCACPYSYKGD 951  
OY 1065 CTACAGTTTTCGCCGAGGACACATTTGA 1091  
Db 952 CTVPINICIONPCQHGGTCHLSDSHKD 978

Search completed: July 5, 2003, 12:55:00  
Job time : 344 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 07:07:38 ; Search time 101 Seconds  
(without alignments)  
6221.591 Million cell updates/sec

Title: US-09-937-521-15  
Perfect score: 2049  
Sequence: 1 ttcaagtgctactactatg.....ttgaagttgctactactag 2049

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.2	5.9	1941	US-09-328-501-2	Sequence 2, Appli
2	121.2	5.9	2013	US-09-328-501-16	Sequence 16, Appli
3	80	3.9	4411529	US-09-103-840A-1	Sequence 1, Appli
4	74.2	3.6	4403765	US-09-103-840A-2	Sequence 2, Appli
5	37.2	1.8	7218	US-08-232-463-14	Sequence 14, Appli
6	36.4	1.8	1716	US-08-656-034-9	Sequence 9, Appli
7	36.4	1.8	2160	US-08-656-034-1	Sequence 1, Appli
8	33.4	1.6	1608	US-09-134-001C-2407	Sequence 2407, Ap
9	33.2	1.6	1882	US-08-257-073-12	Sequence 12, Appli
10	33.2	1.6	1884	US-08-257-073-8	Sequence 8, Appli
11	33	1.6	1554	US-09-134-001C-2607	Sequence 2607, Ap
12	32.6	1.6	2518	US-09-011-745-1	Sequence 1, Appli
13	32.6	1.6	5865	US-09-011-745-8	Sequence 8, Appli
14	32.4	1.6	2172	US-08-030-410-2	Sequence 2, Appli
15	31.8	1.6	1021	US-09-465-558-61	Sequence 61, Appli
16	31.6	1.5	500	US-08-545-809A-37	Sequence 37, Appli
17	31.6	1.5	531	US-08-642-274D-24	Sequence 24, Appli
18	31.4	1.5	531	US-08-952-014C-24	Sequence 24, Appli
19	31.4	1.5	2869	US-08-961-527-130	Sequence 130, Appli
20	31.2	1.5	289	US-09-007-005-17	Sequence 17, Appli
21	31.2	1.5	289	US-09-244-796-17	Sequence 17, Appli
22	31.2	1.5	1140	US-08-896-410-1	Sequence 1, Appli
23	31	1.5	1227	US-08-961-527-250	Sequence 250, App
24	31	1.5	1935	US-08-709-731A-1	Sequence 1, Appli
25	31	1.5	3814	US-09-302-812-5	Sequence 5, Appli
26	31	1.5	3814	US-09-511-477-5	Sequence 5, Appli
27	31	1.5	3814	US-09-511-507-5	Sequence 5, Appli

C	28	31	1.5	26700	1	US-08-472-217-1	Sequence 1, Appli
C	29	31	1.5	26700	2	US-08-488-199-5	Sequence 5, Appli
C	30	31	1.5	26700	3	US-08-760-534A-1	Sequence 1, Appli
C	31	30.8	1.5	2184	1	US-07-815-333A-1	Sequence 1, Appli
C	32	30.8	1.5	2288	1	US-08-290-937B-4	Sequence 4, Appli
C	33	30.8	1.5	2289	1	US-07-838-410-2	Sequence 2, Appli
C	34	30.8	1.5	4086	4	US-09-221-017B-363	Sequence 363, App
C	35	30.2	1.5	773	4	US-08-998-416-385	Sequence 385, App
C	36	30.2	1.5	804	2	US-08-767-026-12	Sequence 12, Appli
C	37	30.2	1.5	2277	1	US-08-676-967-2	Sequence 2, Appli
C	38	30.2	1.5	2277	1	US-08-676-974-2	Sequence 2, Appli
C	39	30.2	1.5	2277	2	US-09-098-487-2	Sequence 2, Appli
C	40	30.2	1.5	3209	1	US-07-803-633A-12	Sequence 12, Appli
C	41	30.2	1.5	9144	4	US-08-556-978B-79	Sequence 79, Appli
C	42	30.2	1.5	36651	4	US-09-738-894A-3	Sequence 3, Appli
C	43	30	1.5	420	4	US-09-134-001C-1576	Sequence 1576, Ap
C	44	30	1.5	2091	1	US-08-765-081-4	Sequence 4, Appli
C	45	30	1.5	2091	3	US-09-098-082-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1									
US-09-328-501-2									
Sequence 2, Application US/09328501A									
Patent No. 6258581									
GENERAL INFORMATION:									
APPLICANT: ITO, Makoto									
APPLICANT: OKINO, No. 6258581omu									
TITLE OF INVENTION: Ceramidase Gene									
FILE REFERENCE: 1422-0377P									
CURRENT APPLICATION NUMBER: US/09/328, 501A									
CURRENT FILING DATE: 1999-06-09									
EARLIER APPLICATION NUMBER: 10-234769 JAPAN									
EARLIER FILING DATE: 1998-08-20									
NUMBER OF SEQ ID NOS: 18									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 2									
LENGTH: 1941									
TYPE: DNA									
ORGANISM: Pseudomonas aeruginosa									
FEATURE:									
OTHER INFORMATION: any n or xaa = unknown									
US-09-328-501-2									
Query Match									
Best Local Similarity 49.7%; Pred. No. 1.8e-30;									
Matches 340; Conservative 0; Mismatches 338; Indels 6; Gaps 1;									
QY	9	CTACTACATTTGGCGTTGGGAGAGCGGATTTGCACAGGCAAGTGCATATATTCATTTGAT	68						
DB	12	CTACCGCTTGGGCTGGGCAAGGCGGACATCCGCGCAAGCCGCAAGTCGGCATGAT	71						
QY	69	GGGCTATAGGCAAAAATGCGAGATGACGGGGCTCTCCACCGAGCGTTCAGCCGTGC	128						
DB	72	GGGTACTCTCTCCCTCGCAAGAGAGACCGCGCATCATGCGCCAGTGGCGGCTGC	131						
QY	129	TTTATCTTGGGCGGATCCAGATGGGTCAAAATCGAATGGCATTTGTGAGCGTGAATG	188						
DB	132	CTTGATGATGAGAGACCGCGGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	191						
QY	189	TATGATTTCCCAAGCACTGAGGTGGAGGCTCTGAAAGACATGAGAGTAAATATGCTC	248						
DB	192	GATGACCTTCCAGGCGCTGCACTGGAAGGCTCTGCGCCGCTCAAGCGGAATACCCCG	251						
QY	249	TCTGTATGGAAGAGCAATGTTATCTGAGTGCATTCACACACACTCTGGCCAGAGC	308						
DB	252	TGCTGTAGAGAGAGCAACAGTGTGCTGCGCCACACACACACTCTGCGTGGCGG	311						
QY	309	GTTTTCATATATATACACTATATATAGTCCAGGAGGATTCAGCAACCGGACCTTTCA	368						
DB	312	CTTCTCCACTAGCGGATGTAACAACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	371						

[illegible]

## RESULT 2

```

1 Sequence 16, Application US/09328501A
2 Patent No. 6258581
3 GENERAL INFORMATION:
4 APPLICANT: OKINO, Gen 6258581omu
5 APPLICANT: ITO, Makoto
6 TITLE OF INVENTION: Ceramidase Gene
7 FILE REFERENCE: 142Z-0377P
8 CURRENT APPLICATION NUMBER: US/09/328, 501A
9 CURRENT FILING DATE: 1999-06-09
10 EARLIER APPLICATION NUMBER: 10-234769 JAPAN
11 EARLIER FILING DATE: 1998-08-20
12 NUMBER OF SEQ ID NOS: 18
13 SOFTWARE: PatentIn Ver. 2.1
14 SEQ ID NO: 16
15 LENGTH: 2013
16 TYPE: DNA
17 ORGANISM: Pseudomonas aeruginosa
18 FEATURE:
19 OTHER INFORMATION: any n or xaa = Unknown
20 US-09-328-501-16

```

[illegible]

QY	309	GTATTTCACATATACACTCTATATATCTGGCCAGCGAGGGATTCAGCAACCGGACCTTTC	368
Db	384	CTTCTCCCACTACGCCATGTACAACTGTGCGTGTGGCTCCAGGAAAAAGACCTTCAA	443
QY	369	GTAACATAGTCTCTGGGATCATGAAAGCATTTGATATAGCTCACAACAACTTTAAAC	428
Db	444	CGCCATCGTTCAGCGGATGTCCGCTCCATCCATCGACGGGGCCAGGCCAGGTTCAGCCGG	503
QY	429	CAAAATCTTTATCACAAGAAAGATGTTGCTAATGTGAGATTCAAACCGAAGCCCTCTC	488
Db	504	CGCCCTGTTCTACGGGAGCGGCGACCTGGCACAACCCAGCCAAACCGTTGCGTGTGTC	563
QY	489	TTACCTTTCGAATCCACAGTCTCAGAGAGAGCAAGTATCTTCAACACAGACAGAGAA	548
Db	564	GCACCTCGAAGAATC-----CGGACATGCGCGGCTACAGAGATGGCATCGACCCGCGAT	617
QY	549	GCTGCTCTGAAACCTGTGGATTTGAATGGAGAACACTTGGGCTTATCAGCTGGTTGC	608
Db	618	GAGCGTGTACAGCTTGTCGACGCCAACGGGAGAGTGGCCGGCGGATCAAGTTGGTCCC	677
QY	609	CATCCACCCCGTBSAGATGAACAATATGCAACACACTTGTGAAATAGTACATATAGGGCTA	668
Db	678	GGTGCACACACACCTCGATATGCAACGCCCAATCTACCTGATCTCCCGGACAACAAGGGCTA	737
QY	669	TGCGGCTTACCTTTTGGACAGA	692
Db	738	CGCCTCTTATCACTGGGAGACAGA	761

### RESULT 3

```

US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007 00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1996-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match	3.9%	Score 80	DB 4	Length 4411529
Best Local Similarity	46.2%	Pred. No. 5e-14		
Matches 380	Conservative 0	Mismatches 430	Indels 12	Gaps 3
QY	25	GGGAGAGCGGATGTGCACAGCAGACAGTGTGCATATCAATTTGATGGCTATGCGCAAAAT	84	
Db	769577	GGCATCGCCGACATCATCTGTGAGGCGGGGCGCTCGGCATATCTGCGTAGCGACATGCC	769518	
QY	85	GGCGCAATATGCACGGGGCTCTCTCACACAGCGATTCACGCGCTTTTATCTTGGCGGA-	143	
Db	769517	GACACACGCAACCCCGGCAATTACCAACGACACTTAGGTACAGCGGCTTGTGTCCGGGAC	769458	
QY	144	-----TCCAGATGGGTCAATCGAATGGCAATTTGTGACGCGTGAACATATGATGATTTCC	198	
Db	769457	GATTTCCACAGACGGGAGCGCGGGTGTGTGCTGATTTGCGCGCAACTGCGCGTCCCATG	769398	
QY	199	CACGACGTGAGTTGGAGGTTCTCGAAGACATAGAGATTAATTTAGGCTCTCTATCGA	258	
Db	769397	CAAAACGTGATGAGGAATGCTGGCGCGGTTGGCGCAATTTGTATGGCAACTACTCG	769338	
QY	259	AGAGACAATGTTATCTCGATGCCAATTACACACACTCTGGCCGACGAGGTTTTCCAA	318	



Db 769337 GAGCAGAAACACACTGATACAGGGGACCCACACCCCGGGCTGCGGTACTGCGGA 769278  
QY 319 TATACACTATATATCTACCGCAGGAGGATTCAGCAACCGGAACTTTCAGTATATGTC 378  
Db 769277 TACCTGCTGTACACCTGACACGAGCGGCTTTCGCGGCAACCTTTCGCGCATGTC 769218  
QY 379 TCTGGATTCATGAAAGACATTTGATATAGCTCACACAATTCCTTAACCGCAAAATCTTT 438  
Db 769217 GAGGAAATTTGCAATGGGGAACAATGCCATGCCATGATGGGCCCGCGAGTTTCG 769158  
QY 439 ATCAACAAGGAATTTGCTTAATGTGCAATCAACCGCAAGCCCTTCCTTTACCTTCTG 498  
Db 769157 CTCTCGACGCGCGAGCTGTAGCGGCAAGCATATGATCCCATGAGCTTTCAGCGTTCAGCGC 769098  
QY 499 AATCCAGTCAGAGAGAGAGATTTCTTCAACACAGACAAAGAAATGCTGCTTG 558  
Db 769097 AACCCGCGCGCGCAAGGGCTTCTCCCAAGCGGCTGACCCGCTACACAGCTGTC 769038  
QY 559 AAACGTGATTTGTAATGGAAGAAGACTTGGGCTTTCAGCTGTTTGCATTCACCGC 618  
Db 769037 ---CGTATGACCGGGCGAGGCGCACCGTGGGCTGATTCATTTCTTCGCTACCGACGCT 768981  
QY 619 GTGAGCATGAACATACCAACCACTTGTGAATAGTGAATATGAGCTATGCGGCTTAC 678  
Db 768980 ACAGCATGACAGATCGCAACCACTCATCTCCGAGACAAAGAGGTTTCGCGGCTAT 768921  
QY 679 CTTTGTGAGCAAGAAAAG---AACAAAGCTATCTGCTGACAGGAGCCGTTGTGCA 735  
Db 768920 CACTGGAAACGACCGCTGGGCGCGGATATCTCCGCGGCTGATTCATTCATGCGC 768861  
QY 736 GGGTTTGCCTTCATCAATCTCGAGAGCTGTCAACCAATTCCTGCGCGCATTTGTC 795  
Db 768860 GCCTTGCACAGAGATTCGCGGCGACATGAGCCGGAATGTCAGCGGCGCTTTCCTCC 768801  
QY 796 AACACAGGGAGTCTTGTGACAAAGACAGACGACCTGTCC 837  
Db 768800 GAGGCCCCACACAGACGCGGATTCGACAAACCCGCTGTAC 768759

RESULT 4  
US-09-103-840A-2/c  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103.840A  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 3.6%; Score 74.2; DB 4; Length 4403765;  
Best Local Similarity 45.8%; Pred. No. 4.8e-12;  
Matches 377; Conservative 0; Mismatches 433; Indels 13; Gaps 3;

QY 25 GGGAGAGCGGATTCACAGACAAGTGTGATATCAATTTGATGGGCTTGGCAAAAT 84  
Db 771538 GGCATGCGCCGACATCACTGTTGAGGCGCGGACTGCGCATGCTCGGCTACGCGCAACTCC 771479

QY 85 GGCCAGATGACGCGGCTCTCTCCACCAGCGTGTTCAGCGGCTTTATCTTGGCGGA- 143  
Db 771478 GACCAGGCGACCGCGGCTTTCACACGACTTATGTCACAGGCGGCTTGTTCGCGGAC 771419  
QY 144 -----TCCAGATGGGTCAATTCGAATGGCATTTGTGAGCGTGAATATGATTTTC 198  
Db 771418 GATTCGCCAGGAGCGGAGCGGCTGCTGCTGATTTGCGCGCAATGCGCGCCCATG 771359  
QY 199 CAACGACTGAGGTGAGGCTCTTAAGAGACTAGAGATTAATATGCTCTGTATCGA 258  
Db 771358 CAAAACGTGATGAGGAAGTGTGCGCGGCTTTCGCGCATTTGATGAGCATATCTTCG 771299  
QY 259 AGACACATGTTATTCAGAGTCCGATTCACACACACTCTGCGCAGAGGCTTTTCCAA 318  
Db 771298 GAGCAGAAACACACTGATCAGCGGACCCACACCCAGCGGCTTTCGCGGCTTTCGCGG 771239  
QY 319 TATACACTATATATCTGCCAGCGAGGATTCGCAACCGGACCTTTCAGTATAGTC 378  
Db 771238 TACCTGCTGTACACCTGACACGAGCGGCTTTCGCGGCACTTTCGCGGCTATGCTC 771179  
QY 379 TCTGGATTCATGAAGACATTTGATATAGCTCACCAATCTTAAACCGCAAAATCTTT 438  
Db 771178 GACGAAATTTGCAATGCGTGAACATGCGCATGCGATGCGCGCGCGGAGGTTTCG 771119  
QY 439 ATCAACAAGGAATTTGCTAATGTGCAATCAACGAAAGCCCTCTTACCTTCTG 498  
Db 771118 CTCTCGACGCGGAGCTGTACGCGGCAAGCATTCATATGATTCCTATAGCTTGCACCGC 771059  
QY 499 AATCCAGTCAGAGAGAGCAAGGATTTCTTCAACACAGACAAAGAAATGCTGCTTG 558  
Db 771058 AACCCGCGCGCGCAAGGGCTTCTCCGAAAGCGTTCGACCGCGATACAGCTGCTG 770999  
QY 559 AAACGTGATTTGTAATGGAAGAGACTTGGGCTTATACAGTGTGTTGCCATTCACCCC 618  
Db 770998 ---CGTATGACCGGGCGAGGCGCACCGTGGGCTGTATTCATTTCTTCGCTACCGAGGT 770942  
QY 619 GTGAGCATGAACATACCAACCACTTGTGAATAGTGAATATGAGCTATGCGGCTTAC 678  
Db 770941 ACAGCATGACGAATTCGCAACCACTCATCTCCGAGACAAAGGTTTCGCGGCTAT 770882  
QY 679 CTTTGTGAGCAAGAAAAGCAAGGCTAT---CTCCGTGACAGGAGCGGCTTGTAC 734  
Db 770881 CACTGGAGACGACCGTGGGCGGCGGATATATCTCCGCGCAAGCCCGCATTCATGCGC 770822  
QY 735 AGCTTGTCTTCATCAATCTCGAGACGCTGTACCCCAATTTTGGCGCATTTGT 794  
Db 770821 CGCTTGCACAGAGCAATTCGCGGCGACATGAGCCCAATGTGACGCGGCGCTTCTCC 770762  
QY 795 CAACACAGGGAGTCTTGTGCAACGACAGAGGACCTGTCC 837  
Db 770761 CGAGGCCCCACACGACCGGGAATTCGACAAACCCGCTGTAC 770719

RESULT 5  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PT2gpt-fls  
US-08-232-463-14

Query Match 1.8%; Score 37.2; DB 1; Length 7218;  
Best Local Similarity 5.5%; Pred. No. 0.19;  
Matches 21; Conservative 194; Mismatches 167; Indels 0; Gaps 0;

1644 GGATGAGCACCACCAATGGCAACATTTGGGGATGCTTGCAGCCAGCAAAACGTGAATA 1703  
DB 1404 RRR 1345  
QY 1704 CAGAGTGGAGAGAGTGTGAAGTATATTTGAGCCGCTAACCAAGAAATTCAGACA 1763  
DB 1344 RRR 1285  
QY 1764 GAACGAGCCCATCAACCTCTCTACTGTGAGAAATACGAGAGCTGTAGCTGACTG 1823  
DB 1284 RRR 1225  
QY 1824 GCAGATATGTATATGATGCTCTCTGGAGAGAGAGCTTTTATTTGACAAAGAAATACT 1883  
DB 1224 RRR 1165  
QY 1884 GGGTCTGAGCAATGCAACATATATCTGCATATTCAGATATCTGCTACCTGAAATCTA 1943  
DB 1164 RRR 1105  
QY 1944 CAGATATGATATTTTGACACCAATTCGAGAGAGAACTTTCGAACCCGCTGTACTACT 2003  
DB 1104 RRR 1045  
QY 2004 AGCATTTGAGAAATTTCTCT 2025  
DB 1044 AGCCAGCTCGAATTAATTTCT 1023

## RESULT 6

US-08-656-034-9  
Sequence 9, Application US/08656034  
Patent No. 6015691 5786172  
GENERAL INFORMATION:  
APPLICANT: Walker, David H.  
APPLICANT: Yu, Xue-Jie  
TITLE OF INVENTION: IMMUNODOMINANT 120 KDA SURFACE-EXPOSED  
ADHESION PROTEIN GENES OF EHRILCHIA

TITLE OF INVENTION: CHAFFEENSIS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/656.034  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Cordeir, Timothy S.  
REGISTRATION NUMBER: 38,414  
REFERENCE/DOCKET NUMBER: UTSG:191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1716 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 57..1700  
US-08-656-034-9

Query Match 1.8%; Score 36.4; DB 3; Length 1716;  
Best Local Similarity 53.5%; Pred. No. 0.12;  
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

633 TAGCAACACCTTGTGAATATGACATATGAGCGGTACCTTTTGGACAGA 692  
DB 149 TAATAGATATTTGTCACACACAGATCTGGGTTCTGAATTTATCAGACAGCAAGA 208  
QY 693 AAAGAAAGGCTATGCTGCTGACAGAGAGCGCTTTGTAGAGCGTTTCTTCATCAA 752  
DB 209 AGATGACACAGCTCTCAACCTTCAATAGAGCAATTTGTGAGAAAGTATTTCTAA 268  
QY 753 TCTCGAGACGCTGCACCCAC 774  
DB 269 AGTTGACACGAAAAAATACTAAC 290

## RESULT 7

US-08-656-034-1  
Sequence 1, Application US/08656034  
Patent No. 6015691 5786172  
GENERAL INFORMATION:  
APPLICANT: Walker, David H.  
APPLICANT: Yu, Xue-Jie  
TITLE OF INVENTION: IMMUNODOMINANT 120 KDA SURFACE-EXPOSED  
ADHESION PROTEIN GENES OF EHRILCHIA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentlin Release #1.0, Version #1.30
5
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/656,034
8 FILING DATE: Concurrently Herewith
9 CLASSIFICATION: 514
10
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Corder, Timothy S.
13 REGISTRATION NUMBER: 38,414
14 REFERENCE/DOCKET NUMBER: UTSG:191
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (512) 418-3000
17 TELEFAX: (512) 474-7577
18
19 INFORMATION FOR SEQ ID NO: 1:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 2160 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: 171..2054
29
30 US-08-656-034-1

```

	Query Match	1.8%;	Score 36.4;	DB 3;	Length 2160;
	Best Local Similarity	53.5%;	Pred. No. 0.15;		
	Matches	76;	Conservative	0;	Mismatches 66; Indels 0; Gaps 0;
OY	633	TAGCAACCATTTGTGTGAATAGTGCACAAATATGGGCTATCGGCTTTACCTTTTGTGACACAGA	692		
Db	263	TAAATAGAAATTTTGTGTGCAACCAACAGAGATCTGGTTCGAATTTATCA;CAAGAGCAGAGA	322		
OY	693	AAAGAACCAAGGCTATCTGCCTGGACACAGGACCGTTTACGAGGCTTTCATCAAA	752		
Db	323	AGATGACACACATCTCTCAACCTTCATTATAGACCATTTGTTCAGAAACGTGAAGTTTCTTAA	382		
OY	753	TCTCGAGAGCTGTACCCACAC	774		
Db	383	AGTTGAACACAGAAAAAACA	404		

RESULT 8  
 US-09-134-001C-2407  
 Sequence 2407, Application US/09134001C  
 Patent No. 6380370  
 GENERAL INFORMATION:  
 APPLICANT: Lynn Doucette-Stamm et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: GTC-007  
 CURRENT APPLICATION NUMBER: US/09/134,001C  
 PRIOR FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/064,964  
 PRIOR FILING DATE: 1997-11-08  
 PRIOR APPLICATION NUMBER: US 60/055,779  
 PRIOR FILING DATE: 1997-08-14  
 NUMBER OF SEQ ID NOS: 5674  
 SEQ ID NO 2407  
 LENGTH: 1608  
 TYPE: DNA  
 ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-2407

Query Match	1.6%	Score 33.4	DB 4	length 1608
Best Local Similarity	48.2%	Pred. No. 1.2		
Matches	94	Conservative	0	Mismatches 101
			Indels	0
			Gaps	0
QY	1586	GTGTCCCGACGACATTCCTTCAAAAATCTAATAGCTTCACTTATTCCTAATATTCGGG	1645	
Db	110	GTGTGGGAAATCTACATCTTTGAAGATTATTAACGGCGAAATGATTCACACGATGGTC	169	

QY 1446 TTAAAGCAACCAATTGGCAAACTTTGGGAGATGCTTCAGCAGCAAAACCTGAATACA 1705  
 Db 170 ATGATATCTTAGTAAGTAAGATGACGGCTTGGCTGTGTTAAACAAGATCATTTTGCTTATG 229  
 QY 1706 GACTGGCGAAGATGGTTGAAGTTATTTGTAGCGCGTAACCCAAAGATTCAGCAGAGA 1765  
 Db 230 AAGATGAACGTCGTTTACATCTTGTGATTAAGGACATGAACGTTTGATCAAGTGATGA 289  
 QY 1766 ACCAGACCCATCAAA 1780  
 Db 290 AAGAGAAAGATGAAA 304

US RESULT 9  
US-08-257-073-12  
Sequence 12, Application US/08257073  
Patent No. 576597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: de Taisne, Charles  
APPLICANT: Tine, John A.  
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtiss, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CRTMS  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1882 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-12

[illegible]

DB 1032 GAGTCTTCGATCAACCTTAACAATATGACCAACATTATTAACAGATTATGAAAAATTAA 1091  
QY 659 ATATGG 664  
DB 1092 AGAAGC 1097

## RESULT 10

US-08-257-073-8  
Sequence 8, Application US/08257073  
Patent No. 5766597

## GENERAL INFORMATION:

APPLICANT: Proletti, Enzo  
APPLICANT: de Taisne, Charles  
APPLICANT: Tine, John A.  
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtiss, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
CLASSIFICATION: 424  
FILING DATE: 09-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1884 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-8

## Query Match

Best Local Similarity 1.6%; Score 33.2; DB 1; Length 1884;  
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 539 ACAAGGAATGCTGCTTGAAGACTGTGATTTGAATGAGAAAGACTTGGCTTATCA 598  
DB 972 ACATGTAATGATTAATTTCCAGCAATTTGATCTTTTAATGATATAATAGTTTGAATT 1031  
QY 599 GCTGTTGGCATCCACCCGCTGAGCATGAACAATAGCAACCACTTTGTGAATAGTACA 658  
DB 1032 GAGTCTTCGATCAACCTTAACAATATGACCAACATTATTAACAGATTATGAAAAATTAA 1091  
QY 659 ATATGG 664

DB 1092 AGAAGC 1097

## RESULT 11

US-09-134-001C-2607  
Sequence 2607, Application US/09134001C  
Patent No. 6380370

## GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2607  
LENGTH: 1554  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2607

## Query Match

Best Local Similarity 1.6%; Score 33; DB 4; Length 1554;  
Matches 111; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 1518 TGCAATACATCCACACTCTTCAGAGACCTTGCTAAGCAATTCGTAAGCAGACAGTAGCCAA 1577  
DB 726 TGAATTCGTGCAACCGTTTATAGTTGAAGCCAAACATGCTGTAACGATGGAGACG 785  
QY 1578 CATGACAGTGTGCTCCGAGCCCTCCATCTTCAAAATCTAATAGCTTACTATTCCTAA 1637  
DB 786 AGTATCTTTTATATTTCCGTCAGATAGACACTGCACTTCAAGAACTTTACTAA 845  
QY 1638 TATTCGGATAGAGCACCATTGGCAAAACATTTGGGATGCTTGCAGCCAGCAAAAC 1697  
DB 846 TAAACGCTTTGATGATTTAAAGTTGAACAAGTGGACAACTATCTACGCTACATTCAC 905  
QY 1698 TGAATACAGAGGAGAGAGTGTGATGATATTTGTAGGCGCTAACCCAAAGATTTC 1757  
DB 906 GAAATATATGACAAATGTAGATGCTGAATTTGTATTTGAAAAAGTTGACTTAATATATAC 965  
QY 1758 A 1758  
DB 966 A 966

## RESULT 12

US-09-011-745-1  
Sequence 1, Application US/09011745  
Patent No. 6165715

## GENERAL INFORMATION:

APPLICANT: Collins, Mary KL  
APPLICANT: Weiss, Robin A  
APPLICANT: Takeuchi, Yasuhiro  
APPLICANT: Cosset, Francois-Loic  
TITLE OF INVENTION: Expression systems  
FILE REFERENCE: 09/011,745  
CURRENT APPLICATION NUMBER: US/09/011,745  
CURRENT FILING DATE: 1998-06-22  
EARLIER APPLICATION NUMBER: PCT/GB96/02061  
EARLIER FILING DATE: 1996-08-23  
EARLIER APPLICATION NUMBER: GB9517263.1  
EARLIER FILING DATE: 1995-08-23  
NUMBER OF SEQ ID NOS: 29  
SEQ ID NO 1  
SOFTWARE: Patentin Ver. 2.0  
LENGTH: 2518  
TYPE: DNA  
ORGANISM: RD114

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)  
OTHER INFORMATION: n is any nucleotide  
US-09-011-745-1

Query Match 1.6%; Score 32.6; DB 4; Length 2518;  
Best Local Similarity 51.0%; Pred. No. 3.2;  
Matches 77; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 911 GGATCATCTATCAGACGCCAGGAGCTGATGCTTGCCTCCAGAGGTGACGGCC 970  
DB 1947 GGACCTTACTAAGCGCAGACAGGAGGAAATTTGTTAGCTTACAGAAAAATCTGTTT 2006  
QY 971 CAGTGTTCGAGCTCACCAGTGGGTGAACATGACAGATGTAGCGCTCAGCTCAATGCCA 1030  
DB 2007 TTATGCTTACAGAGTCAGAGATTTGTGAGAAACAAATATAGAACCTTACAGAGATTTACA 2066  
QY 1031 CACACACAGTGAAGACGTGTAACCTGCCT 1061  
DB 2067 AAAACGCAGGAAAGCCTGGCAGCAACCACT 2097

RESULT 13  
US-09-011-745-8  
Sequence 8, Application US/09011745  
Patent No. 6165715  
GENERAL INFORMATION:  
APPLICANT: Collins, Mary KL  
APPLICANT: Weiss, Rodin A  
APPLICANT: Takeuchi, Yasuhiro  
APPLICANT: Cosset, Francois-Loic  
TITLE OF INVENTION: Expression systems  
FILE REFERENCE: 09/011,745  
CURRENT APPLICATION NUMBER: US/09/011,745  
EARLIER FILING DATE: 1998-06-22  
EARLIER APPLICATION NUMBER: PCT/GB96/02061  
EARLIER FILING DATE: 1996-08-23  
EARLIER APPLICATION NUMBER: GB9517263.1  
EARLIER FILING DATE: 1995-08-23  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 8  
LENGTH: 5865  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Portion of  
OTHER INFORMATION: construct  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3611)  
OTHER INFORMATION: n is any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3612)  
OTHER INFORMATION: n is any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3613)  
OTHER INFORMATION: n is any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3614)  
OTHER INFORMATION: n is any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3799)  
OTHER INFORMATION: n is any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3800)  
OTHER INFORMATION: n is any nucleotide

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3801)  
OTHER INFORMATION: n is any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3802)  
OTHER INFORMATION: n is any nucleotide  
US-09-011-745-8

Query Match 1.6%; Score 32.6; DB 4; Length 5865;  
Best Local Similarity 51.0%; Pred. No. 5.9;  
Matches 77; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 911 GGATCATCTATCAGACGCCAGGAGCTGATGCTTGCCTCCAGAGGTGACGGCC 970  
DB 2805 GGACCTTACTAAGCGCAGACAGGAGGAAATTTGTTAGCTTACAGAAAAATGCTGTTT 2864  
QY 971 CAGTGTTCGAGCTCACCAGTGGGTGAACATGACAGATGTAGCGCTCAGCTCAATGCCA 1030  
DB 2865 TTATGCTTACAGAGTCAGAGATTTGTGAGAAACAAATATAGAACCTTACAGAGATTTACA 2924  
QY 1031 CACACACAGTGAAGACGTGTAACCTGCCT 1061  
DB 2925 AAAACGCAGGAAAGCCTGGCAGCAACCACT 2955

RESULT 14  
US-08-030-410-2  
Sequence 2, Application US/08030410  
Patent No. 6221359  
GENERAL INFORMATION:  
APPLICANT: Komiyama, Atsushi  
APPLICANT: Nakahata, Tatsutoshi  
APPLICANT: Kubo, Tetsuo  
APPLICANT: Tanaka, Ryuhel  
APPLICANT: Kawano, Genji  
APPLICANT: Sudo, Tetsuo  
APPLICANT: Sano, Emiko  
APPLICANT: Kojima, Katsuki  
TITLE OF INVENTION: HEMOPOIETIC STEM CELL MULTIPLIER  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: United States  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/030,410  
FILING DATE: 19930521  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Digilio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8898  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 742-4343  
TELEFAX: (516) 742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2172 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2169  
US-08-030-410-2

Query Match 1.6%; Score 32.4; DB 4; Length 2172;  
Best Local Similarity 60.0%; Pred. No. 3.4;  
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 12 CTACATTGGCGCTTGGGAGAGCGGATTCACAGACAGTGTGCAGATCATCATTTGATGGG 71  
|||||  
DB 1839 CTACACTGTGATGATCACTATGATGCGCCATTACAGAGTGCACATCTCTATATATGSG 1898  
|||||

OY 72 CTATGGCAAAAATGGCCAGAAATGCACGGGG 101  
|||||  
DB 1899 AAATGAGAAATGCAGCCAGCATCATCGAGG 1928  
|||||

RESULT 15  
US-09-465-558-61  
Sequence 61, Application US/09465558  
Patent No. 6436657

GENERAL INFORMATION:

APPLICANT: OROZCO JR, EMIL M.

TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES

FILE REFERENCE: B81322 US NA

CURRENT APPLICATION NUMBER: US/09/465,558

CURRENT FILING DATE: 1999-12-17

EARLIER APPLICATION NUMBER: 60/112,734

EARLIER FILING DATE: 1998-12-18

NUMBER OF SEQ ID NOS: 70

SOFTWARE: Microsoft Office 97

SEQ ID NO 61

LENGTH: 1021

TYPE: DNA

ORGANISM: Trillium aestivum

US-09-465-558-61

Query Match 1.6%; Score 31.8; DB 4; Length 1021;  
Best Local Similarity 47.3%; Pred. No. 3.1;  
Matches 96; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 402 TATAGCTCACACAATCTTAACGACGCAAAATCTTTATCAACAAGAAATGTTGCTAA 461  
|||||

DB 136 TTTAGATCATGCTATGTGAGTGAATGTGGCTCTGTTGAAGCTGGATGTTAATCT 195  
|||||

OY 462 TGTGAGATCAACGAGCCCTCTTACCTTCTGAATCCACAGTCAGAGAGCAAG 521  
|||||

DB 196 TGCTAAGCAATCTCAACACAAAGGTTATGAGTGAATGTTAGTAGCAATCAACAA 255  
|||||

OY 522 GTATTCTTCAACACAGACAGAAATGCTGCTTTGAACCTGTGATTTGATGAGAGA 581  
|||||

DB 256 ATTTCACACAGACAGACAGAGCGTGAATGGAAGTTGTGAAGGCGGCTATGACAGCTGG 315  
|||||

OY 582 AGACTTGGCTTTATCAGCTGT 604  
|||||

DB 316 GGCTTTGATGCTGTCTCTCT 338  
|||||

Search completed: July 5, 2003, 09:10:41  
Job time : 109 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 08:28:18 ; Search time 2979 Seconds

(without alignments)  
11139,494 Million Cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgcactactatg.....ttgaagtgcactactag 2049

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estnu:\*  
4: em\_estnu:\*  
5: em\_estcov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	683.2	33.3	700	BB356080	BB356080 BB356080
2	670.2	32.7	896	BB237061	BB237061 6020227402
3	431	21.0	722	BB553338	BB553338 603193475
4	410.8	20.0	672	AA913512	AA913512 0137h09.s
5	369	18.0	534	BB702763	BB702763 602684640
6	344.4	16.8	813	BB387191	BB387191 601277054

7	322.2	15.7	790	BB459539	BB459539 603200546
8	317.6	15.5	636	BB773189	BB773189 602721925
9	314	15.3	436	BB193951	BB193951 UT-R-CN1
10	302.2	14.7	484	BB288059	BB288059 602383833
11	253.8	12.4	742	BB153673	BB153673 603190594
12	250.8	12.2	312	BB215231	BB215231 BB215231
13	241.6	11.8	281	BB356081	BB356081 BB356081
14	231.4	11.3	806	BB971187	BB971187 601651171
15	229.6	11.2	789	BB531642	BB531642 601230651
16	228.4	11.1	246	BB160656	BB160656 AV160656
17	202.8	9.9	277	BB242940	BB242940 BB242940
18	200.8	9.8	698	BB327376	BB327376 BB327376
19	200.4	9.8	936	BB560354	BB560354 603253902
20	193.2	9.4	269	BB284970	BB284970 AV284970
21	174	8.5	674	BB2018768	BB2018768 RPT-23-3
22	164.6	8.0	716	BB1309181	BB1309181 EST530591
23	157	7.7	451	BB220146	BB220146 VY52H04.T
24	155.8	7.6	745	BB602483	BB602483 HVSMEH09
25	151.4	7.4	299	BB556696	BB556696 BBS56696
26	149.6	7.3	648	BB859350	BB859350 fvs5905.Y
27	143	7.0	674	BB1447071	BB1447071 NF032C05F
28	143	7.0	733	BB646686	BB646686 EST508305
29	142	6.9	601	BB705497	BB705497 SK49G10.Y
30	136.6	6.7	384	BB273133	BB273133 sa017d06
31	136.2	6.6	521	BB273133	BB273133 EST85231
32	135.6	6.6	641	BB1532035	BB1532035 SD03384.5
33	135.2	6.6	436	BB817265	BB817265 BB817265
34	133	6.5	569	BB462326	BB462326 EST324591
35	131.8	6.4	693	BB136355	BB136355 RES0638.5
36	131	6.4	701	BB640328	BB640328 SD11318.5
37	130.4	6.4	619	BB1543512	BB1543512 SD10303.5
38	129.2	6.3	634	BB1098334	BB1098334 IPL 31.B0
39	128.6	6.3	759	BB0850241	BB0850241 QGB12A24
40	127.4	6.2	834	BB049419	BB049419 RPT-24-3
41	124.2	6.1	688	BB001648	BB001648 QG68G21
42	124	6.1	562	BB1534650	BB1534650 SD07582.5
43	118.8	5.8	550	BB323458	BB323458 BJ323458
44	118.8	5.8	720	BB012999	BB012999 QG33H13.Y
45	118.2	5.8	594	BB1639271	BB1639271 SD21820.5

## ALIGNMENTS

RESULT 1  
LOCUS BB356080 700 bp mRNA linear EST 24-OCT-2001  
DEFINITION BB356080 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C03011021 3', mRNA sequence.  
ACCESSION BB356080  
VERSION BB356080.2 GI:16405735  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 700)  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M., and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
On Jul 12, 2000 this sequence version replaced gi:9067908.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222





## source

1. .896  
 /organism="Mus.musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="IMAGE:4162513"  
 /clone\_lib="NCI\_CGAP\_L19"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-Sport6; site: 1; NotI;  
 site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.9 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT  
 236 a 223 c 229 g 208 t

Query-Match  
 Best Local Similarity 93.1%; Pred. No. 2.3e-192;  
 Matches 779; Conservative 0; Mismatches 48; Indels 10; Gaps 7;  
 557 TGAACCTGGTGGATTTGATGAGAGACTTGGGCTCTATCATGCTGTGTCATCCATCCAC 616  
 1 TGAACCTGGTGGATTTGATGAGAGACTTGGGCTCTATCATGCTGTGTCATCCATCCAC 60  
 617 CCGTAGCATGAACAATAGCAACCACTTTGTAATAGTACAAATATGCGCTATGCGGCTT 676  
 61 CCGTAGCATGAACAATAGCAACCACTTTGTAATAGTACAAATATGCGCTATGCGGCTT 119  
 677 ACCCTTTGAGCAAGAAAGCAAGGCTATCTGCGTGGACAGGAGCGCTTTGAGAG 736  
 120 ACCCTTTGAGCAAGAAAGCAAGGCTATCTGCGTGGACAGGAGCGCTTTGAGAG 179  
 737 GCTTTCCTTCATCAATCTCGAGAGCTGTCCACCAATCTTGGCCGCAATGTCGA 796  
 180 GCTTTCCTTCATCAATCTCGAGAGCTGTCCACCAATCTTGGCCGCAATGTCGA 239  
 797 ACACAGGGAGCTCTTGTGACACACAGACAGCACTGTCCCAAGCGTGGCTATGACGT 856  
 240 ACACAGGGAGCTCTTGTGACACACAGACAGCACTGTCCCAAGCGTGGCTATGACGT 299  
 857 GCATGGCCAGCGGACCTGGACAGACATGTTGAGACACACATTTATGAGAGGATCA 916  
 300 GCATGGCCAGCGGACCTGGACAGACATGTTGAGACACACATTTATGAGAGGATCA 359  
 917 TCTATCAGAGGCCAAGAGAGCTGTATGCTCTGCTCCAGAGAGTGACCGGCGGAGTGC 976  
 360 TCTATCAGAGGCCAAGAGAGCTGTATGCTCTGCTCCAGAGAGTGACCGGCGGAGTGC 419  
 977 TTGACAGCTCACCAGTGGGTGAACATGACAGATGTGAGGCTCCAGCTTAATGCCACACA 1036  
 420 TTGACAGCTCACCAGTGGGTGAACATGACAGATGTGAGGCTCCAGCTTAATGCCACACA 479  
 1037 CAGTGAAGACGTGAACCTGCGCTGGGCTACAGTTTGGCCGAGGACATTTATGAG 1096  
 480 CAGTGAAGACGTGAACCTGCGCTGGGCTACAGTTTGGCCGAGGACATTTATGAG 539  
 1097 TTTGGGGCTCAATATTACACAGGAACTACGGAAGGAGATTCATTTGGGACACTTTC 1156  
 540 TTTGGGGCTCAATATTACACAGGAACTACGGAAGGAGATTCATTTGGGACACTTTC 599  
 1157 GGGACCAAGCTTTGGGAAAACA-TCTGAAGAGATTGAGAGTGTGAGAAAACCAACCA 1215  
 600 GGGACCAAGCTTTGGGAAAACA-TCTGAAGAGATTGAGAGTGTGAGAAAACCAACCA 659  
 1216 ATCCGCTTCACAGTGGAGAGC--TGAGATACCAATCTTGGGACCAATATT-GTT 1272  
 660 ATCCGCTTCACAGTGGAGAGCCTGACGAATACCAATCTTGGGACCAATATTGTT 719  
 1273 GATGTCAGAC--TTGTTACCGTTGGGCTTGGCATTAGCTATTCCTGGG--AATT 1328  
 720 GATGTCAGACCTTTTACCGTTGGGCTTGGCATTAGCTATTCCTGGGGAATTA 779  
 1329 AACAAACATGTC-GGGACGAGAGATTCTGAGGCAATTTAAAAAATTTGCACTTT 1384  
 780 ACCACCATGTCGGGGAGGAATATTACGTAGACATTCACACATTTTGGCTAT 836

RESULT 3  
 BI553338 722 bp mRNA linear EST 05-SEP-2001  
 LOCUS 603193475F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5264635 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI553338  
 VERSION BI553338.1 GI:15440650  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 722)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: [cgaps@email.nih.gov](mailto:cgaps@email.nih.gov)  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLM1167 Row: e Column: 20  
 High quality sequence start: 3  
 High quality sequence stop: 718.

## FEATURES

Location/Qualifiers  
 1..722

1. .722  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:5264635"  
 /clone\_lib="NIH\_MGC\_95"  
 /tissue\_type="hippocampus"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pBluescript (modified  
 pBluescript KS+); Site: 1: BamHI; Site: 2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TGTGTTTATTTTATTTT-3',  
 size-selected for average insert size 2.5 kb and  
 normalized to R0F 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, In preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT  
 214 a 154 c 168 g 186 t

Query-Match  
 Best Local Similarity 84.5%; Pred. No. 1.7e-119;  
 Matches 496; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

1463 ACCAGGCTCGCGGTAGAGAGCAATCTACATTTATGAGACCAACCGCTGCGAT 1522  
 47 AACAGGCTCGAGGATATGAGAGCAATCTTATGAGACCGACGATTTATGCTT 106  
 1523 ACATCCAGCTTCAGAGAGCTTGAAGCAATTTGCTAGGACAGACAGTACCAACATCA 1582  
 107 ACATTCAGCTTCAGAGAGCTTGTAGAGCTTGTGCTAGGACAGACAGTACCAACCTGA 166  
 1583 GCACTGCTCCGAGCTTCATTTCAAAAATTAATAGCTTACTTATTTCTAATATTG 1642  
 167 GCAGAGCTCCAGAACCTCCCTTTTCAAAATTAATAGTCAATTAATTCCTAGATTG 226  
 1643 CGATATGAGACCAATTTGGGATGATCTTGGACGACGAGAAACCGGAT 1702  
 227 TGGATATGAGACCAAAAGGCAAACTTTCGGGATGCTCTGACGACGAGAAACGATAT 1762  
 1703 ACAGAGTGGAGAGAGTGTGAGATTATTTATGAGCGCTAACCAAGAAATTCAGCAG 1762

Db 287 ACAGAGTGGGGAGTGTCTGGAAGTTATA-TTGTAGCTCTTAACCCGGAAGATTCTAGTAC 345

QY 1763 AGAAGCAGAGCCATCAAACTTCTCTCACTGTGGAGAAATACGAGACTCTGTAGTACT 1822

Db 346 AAACAGAGCCATTCAGACCTTCTCTCACTGTGGAGAAATATAGGCTACTTCAACATCT 405

QY 1823 GCGAGATATGTATTAACGATGCTCTCTGGAGACGAGGTTTATTTGGACAAAGGATATG 1882

Db 406 GCGAGATAGTGTATTAATGATGCTCTCTGGAGACTGTTTATTTATTTGGACAAAGGACTCC 465

QY 1883 TGGGTCTGAGCAGTGAACAATTAATCTAGTATTCAGATATTCAGATCTCTCTGGAATCT 1942

Db 466 TGGGTCTGAGTATTCAGACAGTGTGATTCGATATTCAGACACTGCCAGCTGTGATCT 525

QY 1943 ACAGATTAAGATATTTTGGACACATCGAGAGCAGGAACTTCTGAACCCGCTGTATAC 2002

Db 526 ACAGATTAAGATATTTTGGACACATCGAGAGCAGGAACTTCTGAACCCGCTGTATAC 585

QY 2003 TACGATTTGAAGAAATTTCTTCTCTTTGGAAGTTCTCACTACTTAG 2049

Db 586 TTTCATTTGAAGGACACTTCCCGCTTTGGAAGTTGTAATTTAG 632

## RESULT 4

AA913512/c 672 bp mRNA linear EST 10-JUN-1998

LOCUS 0137109.81 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone

DEFINITION IMAGE:1525697.3 similar to TR:015913 015913 RANDOM SLUG CDNA25

PROTEIN : mRNA sequence.

AA913512

VERSION AA913512.1 GI:3052904

KEYWORDS EST.

## SOURCE

ORGANISM human.

Cell: g9apbs-remail.nih.gov

Embryo: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

REFERENCE 1 (bases 1 to 672)

AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.

TITLE NC1-CGAP http://www.ncbi.nlm.nih.gov/nclogap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

COMMENT Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Small: g9apbs-remail.nih.gov

This clone is available royalty-free through LNL: contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert length: 791 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amerisham

High quality sequence stop: 470.

## FEATURES

source

1. 672

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1525697"

/clone\_1b="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: p77T30-Pec (Pharmacia) with

a modified polylinker; Site:1: Not I; Site:2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCL-CGAP-CG81) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 176 a 146 c 142 g 208 t

Query Match 20.0%; Score 410.8; DB 9; Length 672;

Best Local Similarity 82.4%; Pred. No. 2.3e-113;

Matches 483; Conservative 0; Mismatches 102; Indels 1; Gaps 1;

QY 1465 CAGCTCAGCGGTACGAGGACGATCTACATCTATGACACACACACCTGTGCTATC 1524

Db 672 CAGCTCGCCGATATGAGGCGCATACGAAATATAGTGGCGGACGACCATATCTGCTTAC 613

QY 1525 ATCCAGCTTTCAGAG-ACCTGGTAAAGCAATTTGCTAGGACAGATAGCCAACTAGG 1583

Db 612 ATTCAGCTTTCAGAGAAACCTTGTAGGCTATTTGCTAGGACAGATAGCCAACTAGG 553

QY 1584 CAGTGTCCCGAGGCTCCATCTTCGAAAAATCAATATGCTTCACTTATTCATATTCG 1643

Db 552 CAGAGTCCAGAACCTCCCTTTTCAAAACATTAATATGTTCCATTAATTCCTAGTATG 493

QY 1644 GGATAGACACCAATTTGGCAAACTTTTGGGATGCTTTGACACGACCAAACTTAATA 1703

Db 492 GGATAGACACCAAAAGGCAAACTTTGGGATGCTTTGACACGACCAAACTTAATA 433

QY 1704 CAGAGTGGGAAAGTGTGTAAGTTATTTATTTAGGCGCTTACCCAAAGATTCAGACAG 1763

Db 432 CAGAGTGGGAAAGTGTGTAAGTTATTTATTTAGGCGCTTACCCAAAGATTCAGACAG 373

QY 1764 GAACAGAGCCATCAAACTTCTCTCACTGTGAGAAATACGAGACTCTGTACTGACTG 1823

Db 372 AAACAGAGCCATTCAGACACTTCTCACTGTGAGAAATATGAGGCTACTTCAACATCTG 313

QY 1824 GCAGATTAATGTATACGATGCTCTCTGGAGACGAGGTTTATTTGGCAAAAGAACTACT 1883

Db 312 GCAGATTAATGTATACGATGCTCTCTGGAGACTGTTTATTTGGCAAAAGGAACTACT 253

QY 1884 GGGTCTGAGCAATCAACAAATATCTGCAATATTCGATATTCGCTACCTGGAACTTA 1943

Db 252 GGGTCTGAGTATCAACAGATGCAATATTCGATATTCGCAACACTGCTGAACTTA 193

QY 1944 CAGATTAAGATATTTTGGACAACTTCGAGAGCAGAACTTCTGAAACCCGCTGTACT 2003

Db 192 CAGATTAAGATATTTTGGACAACTTCGAGAGCAGAACTTCTGAAACCCGCTGTACT 133

QY 2004 AGCATTTGAGGAATTTCTTCTCTTTTGAAGTTGTACTACTTAG 2049

Db 132 TTCATTTGAGGACACTTCCCGCTTTTGAAGTTGTACTACTTAG 87

## RESULT 5

BG702763 534 bp mRNA linear EST 07-MAY-2001

LOCUS 602684640F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4817504.5'

DEFINITION mRNA sequence.

ACCESSION BG702763

VERSION BG702763.1 GI:13974426

KEYWORDS EST.

## SOURCE

ORGANISM human.

Cell: g9apbs-remail.nih.gov

Embryo: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

REFERENCE 1 (bases 1 to 534)

AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.

TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: g9apbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihaki

Toshiluyki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: L1AM10717 row: 0 column: 09

High quality sequence stop: 534.

## FEATURES

source

1. 534

/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_95"  
/issue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: pBluescriptR (modified pluescript KS+); Site:1: BamHI; Site:2: SalI; XhoI (gtcgag); Oligo-dt primed using primer 5'-TTTTTGTGTTTGTGTTT-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/MHGR), National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 154 a 121 c 132 g 127 t  
ORIGIN

Query Match 18.0%; Score 369; DB 12; Length 534;  
Best Local Similarity 84.7%; Pred. No. 1.1e-100;  
Matches 414; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1463 ACCAGCTGACGGGTACGAGCAGCATCTACATCTATGACACACACCTGTCGAT 1522  
DB 46 AACAGGCTGAGGATGAGGAGCATGCATATTATGACCGCAGCATATTATCTGCTT 105  
QY 1523 ACATCCACTCTTCAGAGACTGCTAAGCAATGCTAAGGACAGACAGATGATG 1582  
DB 106 AATTTCAGCTCTTCAGAAACCTTGCTAAGGCTATGCTACGACAGGATAGCCACCTGA 165  
QY 1583 GCAGTGGTCCGAGCCTCCATCTTCATAAATCTAATAGCTTAACTCTAATATG 1642  
DB 166 GCAGAGGTCCAGAACCTCCCTTTTCAACAATTAATAGTCCATTAATCTCAGATATG 225  
QY 1643 CGGATGAGCAGCAATTTGCGCAACATTTGGGATCTCTTGCAGCCAG:AAAACCTGAAT 1702  
DB 226 TGGATGAGCAGCAAAAGGAGCAAACTTGGGATCTCTGAGCCAG:AAAACCTGAAT 285  
QY 1703 ACAGAGTGGAGAGGTTGAGTATATTGTTAGGCTTAACCCAAAGAAATTCAGCG 1762  
DB 286 ACAGAGTGGGAGAGTCTGAGATTAATTTGAGTGTACACCGAAGAAATTCAGTAC 345  
QY 1763 AAGACGAGCCATCAACCTTCTCAGTGTGAGAAATAGAGAGACTGTAGTACT 1822  
DB 346 AAACGAGAGCCATCAGCTTCTCAGTGTGAGAAATAGAGGCTACTCAACATCG 405  
QY 1823 GGCAGATATGTAATACGATGCTCTGAGAGACGAGTTTATTTGGACAAAGAAATAC 1882  
DB 406 GGCAGATATGTAATACGATGCTCTGAGAGACTGTTTATTTGGACAAAGGACTCC 465  
QY 1883 TGGGCTGAGCAATCAATATATAGCATATTCAGATATGCTACCTGAGATCT 1942  
DB 466 TGGGCTGAGCAATCAATATAGCATATTCAGATATTCAGATATTCAGATATTC 525  
QY 1943 ACAGAAATA 1951  
DB 526 ACAGAAATA 534

RESULT 6  
BE387191 813 bp mRNA linear EST 21-JUL-2000  
LOCUS 60127054F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3618180 5',  
DEFINITION mRNA sequence.  
ACCESSION BE387191  
VERSION BE387191.1 GI:9332556  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 813)  
NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE

JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@b-remail.nih.gov  
Tissue Procurement: ATCC/DCID/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNU at: image.lnl.gov  
Plate: LNC4286 row: K column: 13  
High quality sequence start: 70  
High quality sequence stop: 706.  
Location/Qualifiers  
1. 813  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_20"  
/issue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOT7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 230 a 155 c 205 g 223 t  
ORIGIN

Query Match 16.8%; Score 344.4; DB 10; Length 813;  
Best Local Similarity 84.5%; Pred. No. 4.8e-93;  
Matches 387; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 1592 CCGAGCCTCATCTCTCAAAATCTAATGCTTCACTTATCTAATATTCGAGATAGAG 1651  
DB 68 CAGAACCTCCCTTTTCAAAACATTAATGTTCCATTATTCCTATATTTGGATAGAG 127  
QY 1652 CACCAATTTGGCAAACTTTGGGAGTCTTCGACGACGAAACCTGAATACAGAGTGG 1711  
DB 128 CACCAAAAGGACAGAACTTTCGGGAGTCTTCGACGACGAAACCTGAATACAGAGTGG 187  
QY 1712 GAGAGTGGTGAATTAATTTGAGGCGCTAACCCAAAGAAATTCAGAGAGACAGA 1771  
DB 188 GGAAGTGGTGAATTAATTTGAGGCGCTAACCCAAAGAAATTCAGAGAGACAGA 247  
QY 1772 CCCATCAACCTCTCAGTGTGAGAAATAGAGAGACTGTAGTGAATTCGAGATAA 1831  
DB 248 CCCATCAACCTCTCAGTGTGAGAAATAGAGAGACTGTAGTGAATTCGAGATAA 307  
QY 1832 TGTATACAGTGGCTCTGAGAGACAGGTTTATTTGGCAAAAGAAATTCGAGTGA 1891  
DB 308 TGTATACAGTGGCTCTGAGAGACAGGTTTATTTGGCAAAAGAAATTCGAGTGA 367  
QY 1892 GCATGCAACAATATATCTGAGATATTCAGATATTCAGATATTCAGATATTCAGATAT 1951  
DB 368 GATATTTTGGACAAATCGGAAGCAGAACTTCGAAACCCGCTGTACATAGCATTTG 2011  
QY 428 GATATTTTGGACAAATCGGAAGCAGAACTTCGAAACCCGCTGTACATAGCATTTG 487  
QY 2012 AAGGAATTTCTTCTCTTTGAGAGTGTCACTACTAG 2049  
DB 488 AAGGACTTCCCGGCTTTGAGAGTGTCACTACTAG 525

RESULT 7  
B1459539 790 bp mRNA linear EST 21-AUG-2001  
LOCUS 603200546F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:526690 5',  
DEFINITION mRNA sequence.

Accession	Version	KeyWords	Source	Organism
B1459539			human.	Homo sapiens
B1459539.1	GI:15250195	EST.		Euryotia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Reference	Title	Journal	Comment	
1 (bases 1 to 790)	NIH-MGC http://mgc.ncl.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	Contact: Robert Strausberg, Ph.D. Email: cgaabg@ncl.nih.gov Tissue Procurement: Miklos Palokovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshiyuki and Piero Carninci (RIKEN) CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LHAM1672 row: k column: 11 High quality sequence stop: 790.	
Features	source	location/Qualifiers		
	1..790	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="5266590" /clone_id="NIH_MGC_97" /lab_host="DH10B" /note="Organ: testis; Vector: BluescriptR (modified pluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (cgccagaa...); Oligo-dT primed using primer 5'-TTTTTTTCTTTTNN-3', size-selected for average insert size 2.2 kb and normalized to Rof 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH/MGC Library."		
Base Count	188 a	198 c	225 g	178 t
Origin				
Query Match	15.7%	Score 322.2;	DB:13;	Length 790;
Best Local Similarity	79.6%;	Pred. No. 2.7e-86;		
Matches	393; Conservative	0; Mismatches	99; Indels	2; Gaps
OY	1463 ACCAGGGTCACCGGTACGAGCAGCATATCAATTATGAGCACACACCCTGCTGCAT			1522
Dd	297 AACGCCGCACCATATGAGCAGCAGCATGACAATTTATGAGACCCGACCATTAATCTGCTT			356
OY	1523 ACATCCAACTCTTGAGAGACCTCTGCTAAGGAATTTGCTAGGAGACAGTAGCCAAATGA			1582
Dd	357 ACATTCAGCTCTTCAGAAACCTCTGCTAAGGCTATTTGCTAGCGAGACCGTAGCCAACTGA			416
OY	1583 GCATGGTCCCAGACCTCCATCTTCCAATAATCTAATAGCTTACTTATTCCTAATATG			1642
Dd	417 GCAGAGGTCAGAAACCTCTCTTTTCAAACATTAATAGTCCATTAATTCCTAGATAG			476
OY	1643 CGGATAGAGCACCAATTTGGGGAATTTGGGAGTCTTGACAGCAGCAAAACCTGAT			1702
Dd	477 TGATAGAGCACCAAAAGGCAAGAACCTTCGGGGATGCTCTCAGCAGCAAAACTGAT			536
OY	1703 ACAGAGTGGAGAAGTGGTTGAAGTTATATTGTAAGGCGCTAACCCTAAGAAATTCAGAC			1762
Dd	537 ACAGAGTGGGGAATTTCTGAAGTTATATTGTAAGGCGCTAACCCTAAGAAATTCAGAC			596
OY	1763 AGAACGAGACCATCAAACTTCTCTACTGTGAGAAATACGAGAGCTCTGTAGCTGACT			1822
Dd	597 AAAACGAGACCATCAAACTTCTCTACTGTGAGAAATACGAGAGCTCTCAACATCTGT			656
OY	1823 GGCAAGTATGTTATTAAGAGTGCCTCCGGGAGAGAGAGGTTTATTTGGCAAAAGGAATAC			1882
Dd	657 GGCAAGTATGTTTATTAAGAGTGCCTCCGGGAGAGAGGTTTATTTGGCAAAAGGAATAC			716

QY	1883	TGGGCTCAGCAATGCACCAAC--TATACTGCGCAATTCACATATCTGCTACCCCGGAAT	1940
Db	717	TGGGCTCTAGTAATATGCACACAGTGGGATGGCCCTATTTCACACACTGGCCAGCCTTGAAT	776
QY	1941	CTACAGATAATAGAT	1954
Db	777	CTACAGATCAGAT	790
RESULT 8			
Bg773189			
LOCUS			
DEFINITION	Bg773189	636 bp	mRNA
ACCESSION	602721925p1	NIH_MGC_97	Homo sapiens
VERSION	Bg773189		cdna clone IMAGE:4838671 5',
KEYWORDS			
SOURCE	Bg773189.1	GI:14083842	
ORGANISM	EST.		
REFERENCE	human.		
AUTHORS	Homo sapiens		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
COMMENT	1 (bases 1 to 636)		
	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: c9qpb5-remail.nih.gov		
	Tissue Procurement: Miklos Rakovits, M.D., Ph.D.		
	CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki		
	Toshiyuki and Piero Carninci (RIKEN)		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLN at:		
	<a href="http://image.lnl.gov">http://image.lnl.gov</a>		
	Plate: LHAM10773	row: a	column: 08
FEATURES	High quality sequence stop: 636.		
source	Location/Qualifiers		
	1..636		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606		
	/clone="IMAGE:4838671"		
	/clone_id="NIH_MGC_97"		
	/lab_host="DH10B"		
	/note="Organ: testis; Vector: pBluescript (modified		
	pBluescript KS+); Site_1: BamH; Site_2: SalI-XhoI (gtcgag		
	); Oligo-dT primed using primer 5'-TTTTTTTCTTTTIVN-3',		
	size-selected for average insert size 2.2 kb and		
	normalized to R0F 5. This is a primary library enriched		
	for full-length clones and constructed using the		
	Cap-trapper method (Carninci, in preparation). Library		
	constructed by M. Brownstein (NIH/NHGRI, National		
	Institutes of Health). Note: this is a NIH-MGC Library."		
BASE COUNT	183 a	131 c	157 g 165 t
ORIGIN			
Query Match	15.5%;	Score 317.6;	DB 12; Length 636;
Best Local Similarity	85.7%;	Pred. No. 5.8e-85;	
Matches 353;	Conservative 0;	Mismatches 59;	Indels 0; Gaps 0;
QY	1638	TATTGCGGATAGACACCAATTTGGCAACATTTTGGGGATGCTTGGACGCCAGCAAAACC	1697
Db	156	TATTGCGGATAGACACCAAAAGCAGAACTTGGGGATGCTCTGCAGCAGCAAAACC	215
QY	1698	TGAATACAGAGTGGGAGAGTGGTTGAAGTTATATTGTAGCGCTAACCAAGAAATTC	1757
Db	216	TGAATACAGAGTGGGAGAGTGGTTGAAGTTATATTGTAGCTCTAACCGAAGAAATTC	275
QY	1758	AGCAGAGAACCGACCCATCAACCTTCCTACGTGGAGAAATACGAGACCTGTAGC	1817
Db	276	AGTCAAAACCCAGACCCATCAGACCTTCCTACGTGGAGAAATATGAGGCTACTTCAAC	335

```

QY      1818 TGACTGGCAGATATGTTATACGATGCTCTGGAGAGAGGTTTATGGACACAAGG 1877
        ||||||| ||| ||| ||||||| ||| ||| ||||||| ||| ||| ||||||| |||
Db      336 ATCGGCGAGTAGTGTGTAATGATGCCCTCTGGAGACTCGTTTATGGACAGAAGG 395
QY      1878 AATAGTGGCTGAGCAATGACAATATATGCGATATTCAGATATCTACCTGG 1937
        ||| ||||||| ||| ||| ||||||| ||| ||| ||||||| ||| ||| |||||||
Db      396 ACTCCGCGCTGAGTAATGACAAGAGTGAATGGCATATTCAGACACTGCCAGCTGG 455
QY      1938 AATCTACAGATATGATTTGGACACATGGGAAGCAACTCTGTAACCCGCTGT 1997
        ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db      456 AATCTACAGATATGATTTGGACACATGGGAAGCAAGCAATTCGTAAGCCTGTGT 515
QY      1998 CATACAGATTGGAAGCAATTTCTCTCTTTGAAGTGTGCTACTACTTAA 2049
        ||||||| ||||||| ||| ||| ||||||| ||| ||| ||||||| ||| |||
Db      516 CATACATTGTAAGGCACTTCCCGGCTTTGTAAGTGTGTAACATATTAG 567

RESULT 9
LOCUS   B0193951/c 436 bp mRNA linear EST 30-APR-2002
DEFINITION UI-R-CNI-cliz-o-04-0-UI s1 UI-R-CNI Rattus norvegicus cDNA clone
ACCESSION B0193951
VERSION   B0193951
KEYWORDS  B0193951.1 GI:20369502
SOURCE    EST.
ORGANISM  Norway rat.
           Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 436)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
           discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
COMMENT    Contact: Soares, MB
           Program for Rat Gene Discovery and Mapping
           University of Iowa
           451 Eckstein Medical Research Building Iowa City, IA 52242, USA
           Tel: 319 335 8250
           Fax: 319 335 9565
           Email: msoares@blue.weeg.uiowa.edu
           The sequence contained an oligo-dT track that was present in the
           oligonucleotide that was used to prime the synthesis of first
           strand cDNA and therefore this may represent a bonafide poly A
           tail. The sequence tag present in the cDNA between the NotI site
           and the oligo-dT track served to identify it as a clone from the
           normalized fundus library cDNA Library Preparation: M.B. Soares Lab
           Clone distribution: clones will be available through Research
           Genetics (www.resgen.com)
           Seq primer: M13 forward
           POLY-A=yes.

FEATURES
source    Location/Qualifiers
           1..436
           /organism="Rattus norvegicus"
           /strain="Sprague-Dawley"
           /db_xref="taxon:10116"
           /clone="UI-R-CNI-cliz-o-04-0-UI"
           /clone_1lb="UI-R-CNI"
           /dev_stage="adult"
           /note="Vector: pT773D-Pac (Pharmacia)" with a modified
           polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-CNI
           library is a subtracted library derived from the following
           pool of seven normalized rat libraries: normalized rat
           seminal vesicles, normalized rat penis, normalized rat
           bladder, normalized rat cervix, normalized rat brown
           adipose, normalized rat fundus, and normalized rat
           salivary gland. It was constructed according to the
           procedure described by Bonaldo, Lennon & Soares (Genome
           Research Genome 6: 791-806, 1996). For construction of
           the CNI library, plasmid DNA from the pool of normalized

```

libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat unigene Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and MN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries C40 and C41 corresponding to plates R-CA0-AMV through R-CA0-AXS, R-CA0-AZX through R-CA0-BAL, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS, R-CA0-BKE, R-CA0-BKH-H, R-CA0-BKJ-K, R-CA0-BKR through R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA, R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through R-CA0-BOI, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through R-CA1-BDA, R-CA1-BHZ through R-CA1-BJF, R-CA1-BJR, R-CA1-BJT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF, R-CA1-BKI, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through R-CA1-BLS, R-CA1-BLS, R-CA1-BLU-V, R-CA1-BNR, and R-CA1-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries C50s, C70s, CUDs, CW0s, CX0s and normalized libraries C50, C70, CUD, CW0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAX, R-CUDs-CBP through R-CUDs-CCA, R-CW0s-CGB through R-CW0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BPD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BOP, R-CT0-BVN, R-CUD-BUQ through R-CUD-BVL, R-CW0-BVY through R-CW0-BMP, R-CW0-BXN through R-CW0-BXO, R-CX0-BMQ through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library C40 and 4,000 from normalized eye library C41) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVO. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BVOP (7-9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BVOP-AOI through R-BVOP-AOX, and R-BX0-AY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including C40, C20, DA0, DB0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ through R-DB0-BZA, R-DC0-BZI through R-DC0-BZO, R-DD0-CAY through R-DD0-CBA, R-DD0-BZR through R-DD0-CAH, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CNO, that makes up the tracer. The corresponding plates are R-CNO-BXW through R-CNO-BLD, R-CNO-BLE, R-CNO-BLP through R-CNO-BLR, R-CNO-BLT, R-CNO-BLM-X, R-CNO-BMB, and R-CNO-BME through R-CNO-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CNO pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkz-c-06-0-UI, bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CNO clone (corresponding to the address bkz-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel







Query Match	12.2%	Score 250.8	DB 10	Length 312
Best Local Similarity	90.8%	Pred. No. 7.8e-65		
Matches 278:	Conservative	0:	Mismatches 27:	Indels 1:
QY	383	GGATCATGAAGAGCATTGATATAGCTCAGACAAATCTTAAACGAGGCAAAATCTTTATCA	442	
Db	8	GCATCATGAGAGCCTTGACATACCTCAGACGAATCTTACACAGGCCCT-TCCTATGC	66	
QY	443	ACAAAGGAATGTTGCTATGTGAGATTCAAACCCAGCCCTCCTCTTACTTGTGATC	502	
Db	67	ACACAGGCCATCTTGCTTATCTCCAGATCAACCGAGCCCTCCTCTTACTTGTGATC	126	
QY	503	CACAGTCAGAGAGAGCAAGGATATCTTCAACACAGACAAAGAAATCTGGTCTTGAAC	562	
Db	127	CACAGTCAGAGAGAGCAAGGATATCTTCAACACAGACAAAGAAATCTGGTCTTGAAC	186	
QY	563	TGTGTGATTTGATGAGAGAGACTGGGTCTTATACGTGGTTCATCCACCCCTGA	622	
Db	187	TGTGTGATTTGATGAGAGAGACTGGGTCTTATACGTGGTTCATCCACCCCTGA	246	
QY	623	GCATGAACAAATAGACACACTTGTGAAATATGACAAATAGGGTATAGGGCTTACCTT	682	
Db	247	GCATGAACAAATAGACACACTTGTGAAATATGACAAATAGGGTATAGGGCTTACCTT	306	
QY	683	TTGACG 688		
Db	307	TTGACG 312		
RESULT 13				
BB356081				
LOCUS	BB356081	281 bp	mRNA	linear
DEFINITION	BB356081	RIKEN full-length cDNA, adult male corpus striatum Mus.		
ACCESSION	BB356081			
VERSION	BB356081.1	GI:9067909		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 281)			
	Kono, H., Alzawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koye, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.			
TITLE	RIKEN Mouse ESTs (Kono, H., et al.)			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsr.riken.go.jp URL: http://genome-gsc.riken.go.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasakihara, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Thermotranscription and thermotranscription of thermolabile enzymes by trihalos and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh, M., Katsunuma, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,			

FEATURES					
	source	location/Qualifiers			
		1..281			
		/organism="Mus musculus"			
		/strain="C57BL/6J"			
		/db_xref="taxon:10090"			
		/clone="CO30011022"			
		/clone_1fp="RIKEN full-length enriched, adult male corpus striatum"			
		/sex="male"			
		/tissue_type="corpus striatum"			
		/dev_stage="adult"			
		/lab_host="DH10B"			
		/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group In Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGATCCACGACTCTTTTGTTCVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCGCGATTAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"			
BASE COUNT	81 a	72 c	55 g	73 t	
ORIGIN					
Query Match	11.8%; Score 241.6;	DB 10;	Length 281;		
Best Local Similarity	91.4%;	Pred. No. 4.6e-62;			
Matches 256;	Conservative 0;	Mismatches 24;	Indels 0;	Gaps 0;	
OY	411 CACAATCTTAACCCAGCAAAATCTTATCAACAAAGAATGTTGCTATGGCAGAT	470			
Db	2 CACCAATCCAGGACGACCAAAATCTTATCAACAGAGCATGTCTTAATGGCAGAT	61			
OY	471 CAACCGAAGCCCTCTCTTACCTTGGAATCCACAGTCAGAGAGCAAGATTTCTTC	530			
Db	62 CCACCGAACCCCTCTCTCTCACTTCGTAATCCACATTCACAGAGAGCAAGTTTCTTC	121			
OY	531 AAACACAGACAGAAATGCTGTGCTTGAACCTGGTGGATTTGAATGAGAAAGACTTGGG	590			
Db	122 ATACACAGACAGAAATGCTGTCTTGAACCTGGTGGATTTGAATGAGAAAGACTTGGT	181			
OY	591 TCTATACAGCGTTTGGCATCCACCCTGAGCATGAACAATGACCAACCACTTTGTGAA	650			
Db	182 TCTATACGCTGGTTTTCCCATCCACCCTGAGCATGAACAATGACCAACCACTTTGTGAA	241			
OY	651 TAGTGACAAATGCGCTATGCGGCTTACCTTTTGAGCAA	690			
Db	242 TAGTGACAAATGCGCTATGCGGCTTACCTTTTGAGCAA	281			
RESULT 14					
BE971187	806 bp	mRNA	linear	EST 04-OCT-2000	
LOCUS	601651171f1 NIH_MGC_81 Homo sapiens	cDNA clone IMAGE:3934649 5'			
DEFINITION	mRNA sequence.				
ACCESSION	BE971187				
VERSION	BE971187.1	GI:10584523			



KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 806)  
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: CLONETECH Laboratories, Inc.  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LINC775 row: m column: 18  
High quality sequence stop: 426.

FEATURES  
source  
1. 806  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3934649"  
/clone\_1lb="NIH-MGC\_81"  
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site\_1: SfilI (ggccctcggcc); Site\_2: SfilI (ggccatctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTAGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCCGCGACATG-dt(30)BT-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."  
BASE COUNT 222 a 216 c 187 g 181 t  
ORIGIN

Query Match 11.3%; Score 231.4; DB 12: Length 806;  
Best Local Similarity 71.7%; Pred. No. 1.4e-58;  
Matches 428; Conservative 0; Mismatches 156; Indels 13; Gaps 9;

1463 ACCAGGCTCAGGGGACGAGGAGATCTACAACTCTATGACACACACCTGTGCAAT 1522  
1523 ACATGCACTCTTCAGAGACCTTGTGAAGCAATTGCTACGACACAGTACGCA-ACATG 1581  
85 ACATTCAGCTCTTCAGAAACCTTGTGAAGCTATTTGCTACGACACGATGACGCGCTG 144  
1582 AGCAGTGTCCGACCTCCATCTTCAAAAATTAATAGCTTCACTTATCTATAT 1641  
145 AGCAGAGGHCAGAACCTCCCTTTTCAAAATTAATAGTTCATTAATCTCTAGTAT 204  
1642 GGGGATAGACCAATTTGGGATTTTGGAGATTTTGCAGCCAGCAAAACCTGAA 1701  
205 GTGGATAGAGCAACCAAGGAGAACTTTCGGGATGTCTTCAGCCACCAAAACCTGAA 264  
1702 TACAGATGGAGAGTGTGAAGTATAT-----TTGAGGCGCTAACCCAAAGATTC 1757  
265 TACAGATGGAGAGTGTGAAGTATATTTGCTAGAGTGTCTAACCCGAGCAATTC 324  
1758 AGCAGAGAACGACCATCAAACTTCTCTACTGTGAGAA-ATACGAGACTCTGTAG 1816  
325 AGTAAAGACAGCA-CCATCAGACCTCTCTACTGTGAGAGACCTATCTACTTCAA 383  
1817 CTGACTGTGAGATATGT-ATACAGATGCTCTCTGGGAGACGAGTTTATGGC-ACAA 1874  
384 CATCGCGGAGATGATGTGATATGATGCTCTCTGGAGCAGACTGATCATTTGGCGACCA 443

1875 AGGAATACCTGGGTCTGAGCAATG--CAACAATATCTGCATATTCAGATACCTGCTAC 1932  
444 GGGACTCCCTGGGCTACGATATGCGACAGTGGAGCATGGCTATTCACACCTGCCAG 503  
1933 CCTGAACTTACAGATTAATATTTTGGACACAACTGAGACGAGACTTGAACCC 1992  
504 C-TGGAACTTACAGCATCATGATATCTGAAGCATGCGAAGCAGACATATCGTGAAGCT 562  
1993 GCTGT-CATACAGCATTTGAGAGAAATTTCTCTCTTTGAAAGTTGTACACTTA 2048  
563 GAGGTGATCATCTTGATTCGAAGAACTTCCCGCGGTGTGAGTTGAGCTATTA 619

RESULT 15  
LOCUS BE531642 789 bp mRNA linear EST 09-ANG-2000  
DEFINITION 601230651F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3594612 5', mRNA sequence.  
ACCESSION BE531642  
VERSION BE531642.1 GI:9760287  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mhc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LINC8769 row: e column: 13  
High quality sequence stop: 587.

FEATURES  
source  
1. 789  
Location/Qualifiers

/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3594612"  
/clone\_1lb="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"  
BASE COUNT 235 a 180 c 225 g 149 t  
ORIGIN

Query Match 11.2%; Score 229.6; DB 10: Length 789;  
Best Local Similarity 77.7%; Pred. No. 4.7e-58;  
Matches 502; Conservative 0; Mismatches 104; Indels 40; Gaps 17;

442 AACAAAGAAATGCTGTAATGAGATGCA-CCGAAACCCCTCTCTTACTCTGAA 500  
71 AGCAAGAAATGCTGTAATGAGATGCAAGGCGAAGCCCTCTCTTACTCTGATGA 130  
501 TCCTCA-----CAGTCAGAGAGAGCAAGTATTTCTTCAACACAGA--CAAGAAATG 549  
131 ATCCGACAGATCAGAGAGACAGACAGCAAGCTATTTCTTCACAAACAGCAAGGGAATG 190  
550 CTGGTCTTGAAGT---GGTGAATTTGAA-----TGAGAGAACTTGGGTCTTATCA---G 599  
191 CTGGTCTTGAAGTCTGGGTGATTTGATATGATGAGAAAGCTGGGTCTTATCAGCG 250

```
OY 600 CTGTTTGCCATCCACCCCG-TGACATGAACAATAGCA-ACCACTTGTGATA-GTGA 656
Db 251 TGGTTCTGCCATCCACCCCGATGAGCATGAACAGTAGACACCACTTGTGATAGTGA 310
OY 657 CAATATGGGCTATGCGGCTTACCTTTTGACGAGAAAAGCAAGGCTAT---CTGC 712
Db 311 CAATATGGGCTATGCGGCTTACCTTTTGACGAGAAAAGCAAGGCGTAGTCTGCC 370
OY 713 CTGGACAGGACCGT-TTGTAGCAGGCTTGTCTTCATCAAACTCGAGAGCTGTACCC 771
Db 371 GTGGACAGGACCGGTTGTTGATGACAGGCTTGTCTTCATCAAACTCGAGAGCTGTACCC 430
OY 772 AACATTTTGGCCCCGATTTGTCAACACAGGGGAGTCTTGTGACA--CGACAAGAGCA 829
Db 431 AACATGCTTGGCCCCGATTTGTCAACAGGGGAGTCTTGTGACAAGCAGACAAGAGCA 490
OY 830 CCTGTCCCAACGG--TGCGCTAGCATGTGCATGGCCAGCGGACCTGGACAAGACATGT 887
Db 491 CCTGTCCCAACGGGTGGGCTAGACATGTGCATGGCCAGCGGACCTGGACAAGACATGT 550
OY 888 TGAGAGCACACACA--TTATAGAGCGATCATCTATCA-GAAGGCCAAGAGCTGTATGC 944
Db 551 TGAGAGCACACACAGTTATAGAGCGATCATCTATCAAGAGGGCCAAAGAGCTGTATTG 610
OY 945 CTCTGCTTCCCA-GGAGGTGACCGGCCAGTGTTCAGCTCACCAGTGGGTGAAC--AT 1001
Db 611 CTCTGACTCCGAGGAGGTGAACGCGCAGAGTAGCAACTCACCGGGGTGTGACCAT 670
OY 1002 GACAGATGTGAGCTCCAGCTCAATGCCACACACACATGGAAGAG 1047
Db 671 GAAAGATGTAGGCTACACAGTCAAGGCCACCCACAGTGAAGAG 716
```

Search completed: July 5, 2003, 10:00:28  
Job time : 2984 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 07:05:53 ; Search time 468 Seconds  
(without alignment)  
9859.711 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgcactacattggtg.....ttgaagtgtcactactag 2049

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

N.Geneseq-101002:\*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	100.0	2049	21	AAA97634	Mouse mature neutr
2	2049	100.0	2271	21	AAA97633	Mouse neutral/alka
3	2049	100.0	3108	21	AAA97632	Mouse neutral/alka
4	2047.4	99.9	4835	21	AAA97635	Mouse neutral/alka
5	1830	89.3	2283	22	AA99977	DNA encoding rat p
6	1511.4	73.8	2507	22	AA22868	Human ceramidase C
7	524.8	25.6	551	22	AAH98033	Murine 7-transmembr
8	441.4	21.5	3058	24	AB199947	Human sialyl trans
9	431.8	21.1	1194	20	AA206239	Human secreted pro

10	396.4	19.3	539	22	AA99979 PCR product isolat
11	334.8	16.3	2553	23	AB102671 Drosophila melanog
12	166.4	8.1	299	21	AA03574 Human secreted pro
13	155.2	7.6	7083	23	AB102670 Drosophila melanog
14	121.2	5.9	1941	21	AA290566 P. aeruginosa matu
15	121.2	5.9	2013	21	AA290573 Mycobacterium tube
16	80	3.8	4411529	22	AA199682 Mycobacterium tube
17	77.6	3.8	565	22	AA199682 Mycobacterium tube
18	74.2	3.6	4403765	22	AA199683 Mycobacterium tube
19	65	3.2	65	24	ABN3673 Mouse spliced tran
20	47.2	2.3	60	24	ABN3675 Human spliced tran
21	40.4	2.0	14041	22	AAH48024 Internal control B
22	38	1.9	38	21	AA97627 Mouse neutral/alka
23	37.4	1.8	462	24	AB165747 Lung cancer relate
24	37.2	1.8	5535	21	AA70184 Plasmodium falcipa
25	36.4	1.8	1920	19	AAV03655 Ethlichia chaffeen
26	36.2	1.8	2253	23	AA584963 DNA encoding novel
27	35.4	1.7	2258	23	AB104267 Drosophila melanog
28	35.4	1.7	4324	23	AB104266 Drosophila melanog
29	35.4	1.7	5460	24	ABK51683 Human ABCG5 upstre
30	34.8	1.7	38	24	AA22672 Human ceramidase c
31	34.8	1.7	1716	19	AAV03656 Ethlichia chaffeen
32	34.8	1.7	2041	23	AA583709 DNA encoding novel
33	34.8	1.7	2998	23	AB109036 Drosophila melanog
34	34.8	1.7	3293	23	AB109000 Drosophila melanog
35	34.6	1.7	10732	21	AA10594 Gene encoding a su
36	34.4	1.7	4590	22	AAH24065 Yeast AOB604-asso
37	34.2	1.7	654	20	AA61836 B. burgdorferi ant
38	34.2	1.7	657	20	AA61836 B. burgdorferi ant
39	34.2	1.7	717	20	AA61691 B. burgdorferi ant
40	34.2	1.7	783	20	AA61835 Borrelia burgdorfe
41	34.2	1.7	910715	20	AA620248 Dog PTEN/MMAC1 Exo
42	34	1.7	582	21	AA98442 Dog PTEN/MMAC1 Int
43	34	1.7	701	21	AA98490 Dog PTEN/MMAC1 Int
44	34	1.7	1317	23	ABV22520 Human prostate exp
45	34	1.7	1317	23	ABV28335 Human prostate exp

#### ALIGNMENTS

RESULT 1	AAA97634	standard; DNA: 2049 BP.
ID	AAA97634	
AC	AAA97634:	
XX		
DT	02-FEB-2001 (first entry)	
XX		
DE	Mouse mature neutral/alkaline ceramidase-encoding DNA, SEQ ID NO:15.	
XX		
KW	Neutral/alkaline ceramidase; mouse; murine; recombinant production;	
KW	cellular ceramidase content control; antibody; membrane synthesis;	
KW	lipid engineering; ceramide metabolism; drug development;	
KW	mature ceramidase; ds.	
XX		
OS	Mus sp.	
XX		
PN	W0200058448-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	24-MAR-2000; 2000MO-JF01802.	
XX		
PR	26-MAR-1999; 99JP-0084743.	
XX		
PA	(TAKI ) TAKARA SHUZO CO LTD.	
XX		
PI	Ito M.	
XX		
DR	WPI: 2000-619079/59.	
XX	P-PSDB; AAB23292.	
XX		

PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
PT reagent for studying structure and functions of ceramide as well as  
PT developing drugs for diseases associated with ceramide metabolism  
XX  
PS Claim 1; Page 65-66; 76pp; Japanese.

CC The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents DNA encoding the  
CC mature mouse neutral/alkaline ceramidase.

XX  
SQ Sequence 2049 BP; 589 A; 470 C; 490 G; 500 T; 0 other:

Query Match 100.0%; Score 2049; DB 21; Length 2049;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGAGTGGCTACTACTGCTGGGAGAGCGGATTCACAGACAAAGTCTCATATC 60  
DB 1 TTGAGTGGCTACTACTGCTGGGAGAGCGGATTCACAGACAAAGTCTCATATC 60  
OY 61 AATTGATGGCTATGCGCAAAATGCGCAGATGCGGAGTCTCTACAGAGCTGTC 120  
DB 61 AATTGATGGCTATGCGCAAAATGCGCAGATGCGGAGTCTCTACAGAGCTGTC 120  
OY 121 AGCGGTCTTTTATCTTGGGGATCCAGATGGGTCAATGCAATGCAATTTGAGCGTG 180  
DB 121 AGCGGTCTTTTATCTTGGGGATCCAGATGGGTCAATGCAATGCAATTTGAGCGTG 180  
OY 181 GAACATATGATGATTTCCCAACGACGTGAGGTGCGTGAAGAGCACTAGAGTAAA 240  
DB 181 GAACATATGATGATTTCCCAACGACGTGAGGTGCGTGAAGAGCACTAGAGTAAA 240  
OY 241 TATGCTCTCTGTATCGAAGAGACATGTATCTGATGTCATTCACACACTCTGCG 300  
DB 241 TATGCTCTCTGTATCGAAGAGACATGTATCTGATGTCATTCACACACTCTGCG 300  
OY 301 CCAGAGAGGTTTTCATATACACTCTATATCTGCGCAGGAGGATTCAGCAACCGG 360  
DB 301 CCAGAGAGGTTTTCATATACACTCTATATCTGCGCAGGAGGATTCAGCAACCGG 360  
OY 361 ACCTTCAGTACATAGTCTCTGGGATCATGAAAGCATGTATATAGCTCACACAAATCTT 420  
DB 361 ACCTTCAGTACATAGTCTCTGGGATCATGAAAGCATGTATATAGCTCACACAAATCTT 420  
OY 421 AAACGAGGCAAAATCTTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 480  
DB 421 AAACGAGGCAAAATCTTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 480  
OY 481 CCCTCTCTTACCTTCTGGAATCCACAGTCAGAGAGACAGATTTCTTCAACAGAGAC 540  
DB 481 CCCTCTCTTACCTTCTGGAATCCACAGTCAGAGAGACAGATTTCTTCAACAGAGAC 540  
OY 541 AAGGAATGCTGCTGCTGAAACTGTGATTTGAATGAGAAAGCTTGGGCTTATAGC 600  
DB 541 AAGGAATGCTGCTGCTGAAACTGTGATTTGAATGAGAAAGCTTGGGCTTATAGC 600  
OY 601 TGGTTTGCATCCACCCGTCGAGCATGAACATAGCAACCTTGTGAATAGGACAAT 660  
DB 601 TGGTTTGCATCCACCCGTCGAGCATGAACATAGCAACCTTGTGAATAGGACAAT 660  
OY 661 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAAGCAAGGCTATCTGCGCGAGAG 720  
DB 661 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAAGCAAGGCTATCTGCGCGAGAG 720  
OY 721 GGACCGTTTGTAGAGGCTTGTCTTCAAAATCTCGAGACGTGTCACCCAAATCTT 780

DB 721 GGACCGTTTGTAGAGGCTTGTCTTCAAAATCTCGAGACGTGTCACCCAAATCTT 780  
OY 781 GGCCCGATTTGTCAACACAGGAGGAGTCTTGTACAAACAGCAAGACACTGTCCCAAC 840  
DB 781 GGCCCGATTTGTCAACACAGGAGGAGTCTTGTACAAACAGCAAGACACTGTCCCAAC 840  
OY 841 GGTGGGCTTACATGTCATGTCAGGAGGAGTCTGACAAAGCATGTTTGAAGAGCAAC 900  
DB 841 GGTGGGCTTACATGTCATGTCAGGAGGAGTCTGACAAAGCATGTTTGAAGAGCAAC 900  
OY 901 ATTATGAGGAGTATCTATACAAAGGCAAGAGGTATGTCCTGCTCCAGAGAG 960  
DB 901 ATTATGAGGAGTATCTATACAAAGGCAAGAGGTATGTCCTGCTCCAGAGAG 960  
OY 961 GTGACCGGCGGAGTCTTGTGACAGTCCAGAGTGGTGAACATGACAGATGTGAGCTCAG 1020  
DB 961 GTGACCGGCGGAGTCTTGTGACAGTCCAGAGTGGTGAACATGACAGATGTGAGCTCAG 1020  
OY 1021 CTCAATGCGCACACACAGATGAGAGGTATTAACCTGCTGGCTACAGTTTCCGCA 1080  
DB 1021 CTCAATGCGCACACACAGATGAGAGGTATTAACCTGCTGGCTACAGTTTCCGCA 1080  
OY 1081 GGCAATATGATGAGTGTGGGCTCAATATTTACAGAGGAACTACGGAAGGGAGTCCA 1140  
DB 1081 GGCAATATGATGAGTGTGGGCTCAATATTTACAGAGGAACTACGGAAGGGAGTCCA 1140  
OY 1141 TTTGGGACACTCTTGGGAGACAGCTCTTGGGAAACATCTGAAGAGATTTAGAGTGT 1200  
DB 1141 TTTGGGACACTCTTGGGAGACAGCTCTTGGGAAACATCTGAAGAGATTTAGAGTGT 1200  
OY 1201 CAGAAACCCAAACCAATCTCTTACAGTGGAGAGAGTGCAGATACCAATCTTGGCAA 1260  
DB 1201 CAGAAACCCAAACCAATCTCTTACAGTGGAGAGAGTGCAGATACCAATCTTGGCAA 1260  
OY 1261 CCAGATATGTTGATGTTGATGATGTTACAGATGTTACCGTGGCTGAGCTAGCTACCT 1320  
DB 1261 CCAGATATGTTGATGTTGATGATGTTACAGATGTTACCGTGGCTGAGCTAGCTACCT 1320  
OY 1321 GGGGAATTAACAAACATGTGGGAGAGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380  
DB 1321 GGGGAATTAACAAACATGTGGGAGAGAGATTTCTGAGGCAATTAAGAAATTTGCA 1380  
OY 1381 CTTTATGAGGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
DB 1381 CTTTATGAGGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
OY 1441 TACATTAACACATATGAAGATACAGAGCTCAGGCTAGCAGAGCATCTACATCTAT 1500  
DB 1441 TACATTAACACATATGAAGATACAGAGCTCAGGCTAGCAGAGCATCTACATCTAT 1500  
OY 1501 GGACCAACACACCTGTCTGCATACATCAACTCTTCAAGAGCACTTGAAGCAATGCT 1560  
DB 1501 GGACCAACACACCTGTCTGCATACATCAACTCTTCAAGAGCACTTGAAGCAATGCT 1560  
OY 1561 ACGGACACAGTACGACATGAGAGAGTGGTCCGAGGCTCCTCATCTTCAAAATCTATA 1620  
DB 1561 ACGGACACAGTACGACATGAGAGAGTGGTCCGAGGCTCCTCATCTTCAAAATCTATA 1620  
OY 1621 GCTTCACTTATTTCTTATATTTGCGATGAGACACATTTGGCAAACTTTTGGGAGTGC 1680  
DB 1621 GCTTCACTTATTTCTTATATTTGCGATGAGACACATTTGGCAAACTTTTGGGAGTGC 1680  
OY 1681 TTTGAGCCAGCAAAACCTGAATACAGAGTGGGAGAGTGTGAAGTTAATTTGTAGGC 1740  
DB 1681 TTTGAGCCAGCAAAACCTGAATACAGAGTGGGAGAGTGTGAAGTTAATTTGTAGGC 1740  
OY 1741 GCTAACCCAAAGATTTAGCAGAGACAGACCCATTAACCTTCTGACAGTGGAGAAA 1800  
DB 1741 GCTAACCCAAAGATTTAGCAGAGACAGACCCATTAACCTTCTGACAGTGGAGAAA 1800  
OY 1801 TACAGAGACTCTGTAGCTGACTGAGATATGATTAACGATGCTCTGGAGACAGG 1860

Db 1801 TACGAGACTCTGTAGCTGACTGCGACATATATGATAACGATTCCTCTGGAGACGAGG 1860  
Qy 1861 TTTATTGGCACAAGGAATATCTGGTGTCTGAGCAATGCAACATATATCTGGCATATTTCCA 1920  
Db 1861 TTTATTGGCACAAGGAATATCTGGTGTCTGAGCAATGCAACATATATCTGGCATATTTCCA 1920  
Qy 1921 GATACCTGCTACCTCTGGAATCTACAGATAATATTTTGGACACAAATCGGAGACAGAA 1980  
Db 1921 GATACCTGCTACCTCTGGAATCTACAGATAATATTTTGGACACAAATCGGAGACAGAA 1980  
Qy 1981 CTCTGGAACCCGCTGTCTATCTAGCATTTGAGAAATTTCTCTTTTGAAGTTGTC 2040  
Db 1981 CTCTGGAACCCGCTGTCTATCTAGCATTTGAGAAATTTCTCTTTTGAAGTTGTC 2040  
Qy 2041 ACTACTTAG 2049  
Db 2041 ACTACTTAG 2049  
RESULT 2  
ID AAA97633 standard; DNA; 2271 BP.  
XX AAA97633:  
XX  
XX  
XX 02-FEB-2001 (first entry)  
XX  
XX Mouse neutral/alkaline ceramidase ORF, SEQ ID NO:1:  
XX  
XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;  
KW cellular ceramide content control; antibody; membrane synthesis;  
KW lipid engineering; ceramide metabolism; drug development;  
KM open reading frame; ORF; ds.  
XX  
XX Mus sp.  
OS  
XX WO200058448-A1.  
XX  
XX 05-OCT-2000.  
XX  
XX 24-MAR-2000; 2000WO-JP01802.  
XX  
XX 26-MAR-1999; 99JP-0084743.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX Ito M;  
XX  
XX WPI: 2000-619079/59.  
XX P-PSDB; AAB23291.  
XX  
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
PT reagent for studying structure and functions of ceramide as well as  
PT developing drugs for diseases associated with ceramide metabolism  
XX  
XX Example 4: Page 57-58; 76pp; Japanese.  
XX  
XX The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents the mouse  
CC neutral/alkaline ceramidase open reading frame (ORF).  
XX  
XX Sequence 2271 BP; 647 A; 541 C; 525 G; 558 T; 0 other;  
Qy  
Query Match 100.0%; Score 2049; DB 21; Length 2271;  
Best local similarity 100.0%; Pred. No. 0;

Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TTCAGTGGCTACTCTATTTGGCGTGGAGACGGATTTGCACAGACAGTGTGATATC 60  
Db 223 TTCAGTGGCTACTCTATTTGGCGTGGAGACGGATTTGCACAGACAGTGTGATATC 282  
Qy 61 AATTTGATGGGCTATGCGAAAAATATGCGCAGAAATCAGCGGGTCTCTACACAGGCTTTC 120  
Db 283 AATTTGATGGGCTATGCGAAAAATATGCGCAGAAATCAGCGGGTCTCTCTACACAGGCTTTC 342  
Qy 121 AGCCGTGCTTTTATCTTGGCGGATTCAGATGGTGTCAATATGCAATTTTGTAGCGTG 180  
Db 343 AGCCGTGCTTTTATCTTGGCGGATTCAGATGGTGTCAATATGCAATTTTGTAGCGTG 402  
Qy 181 GAACATGATATGATTTCCCAACGACTAGGTTGAGGTCCTGAGACACTAGAGATAA 240  
Db 403 GAACATGATATGATTTCCCAACGACTAGGTTGAGGTCCTGAGAGACTAGAGATAA 462  
Qy 241 TATGCTCTCTGTATGCAAGACAAATGTTATCTGATGCTTCAACACACTCTGCG 300  
Db 463 TATGCTCTCTGTATGCAAGACAAATGTTATCTGATGCTTCAACACACTCTGCG 522  
Qy 301 CCAGCAGGCTTTTCCAAATATACACTATATCTCGCCAGCGAGGATTCAGACACGG 360  
Db 523 CCAGCAGGCTTTTCCAAATATACACTATATCTCGCCAGCGAGGATTCAGACACGG 582  
Qy 361 ACCTTTCAATCATAGTCTCTGGGATCATGAAGCAATGTATATAGCTCACAAATCTT 420  
Db 583 ACCTTTCAATCATAGTCTCTGGGATCATGAAGCAATGTATATAGCTCACAAATCTT 642  
Qy 421 AAACAGGCAAAATCTTATCAACAAAGAAATTTCTAATGTGACATCAACGAAAGC 480  
Db 643 AAACAGGCAAAATCTTATCAACAAAGAAATTTCTAATGTGACATCAACGAAAGC 702  
Qy 481 CCTCTCTTACTCTTGAATTCACACTCAGACAGACCAAGATTTCTTAAACACAGAC 540  
Db 703 CCTCTCTTACTCTTGAATTCACACTCAGACAGACCAAGATTTCTTAAACACAGAC 762  
Qy 541 AAGGAATGCTGGTCTTGAACCTGTGATTTGAATGAGAAAGCTTGGCTTATACGC 600  
Db 763 AAGGAATGCTGGTCTTGAACCTGTGATTTGAATGAGAAAGCTTGGCTTATACGC 822  
Qy 601 TGGTTTGCATCCACCCCGTAGCATGAACAAATGACCAACCTTGTGAATAGTACAT 660  
Db 823 TGGTTTGCATCCACCCCGTAGCATGAACAAATGACCAACCTTGTGAATAGTACAT 882  
Qy 661 ATGGCTATGGCGCTTACTCTTGAACAGAAAAAGCAAAAGGCTATGCTGAGACAG 720  
Db 883 ATGGCTATGGCGCTTACTCTTGAACAGAAAAAGCAAAAGGCTATGCTGAGACAG 942  
Qy 721 GGACCGTTTGTAGAGGCTTGTCTTCAATCAATCTCGAGACGTGTACCCACATTTCTT 780  
Db 943 GGACCGTTTGTAGAGGCTTGTCTTCAATCAATCTCGAGACGTGTACCCACATTTCTT 1002  
Qy 781 GGCCCGATTTGTCAACACAGGGGAGCTTGTGACAAAGCAAGACGACACCTGTCCCAAC 840  
Db 1003 GGCCCGATTTGTCAACACAGGGGAGCTTGTGACAAAGCAAGACGACACCTGTCCCAAC 1062  
Qy 841 GGTGGGCTATGACATGTGATGCGCAGCGGACCTGAGACAGCATGTTTGAAGACACAC 900  
Db 1063 GGTGGGCTATGACATGTGATGCGCAGCGGACCTGAGACAGCATGTTTGAAGACACAC 1122  
Qy 901 ATTTATGAGCGATATCTATCAAGAGCCCAAGAGCTGTATGCTCTGCTCCACAGAG 960  
Db 1123 ATTTATGAGCGATATCTATCAAGAGCCCAAGAGCTGTATGCTCTGCTCCACAGAG 1182  
Qy 961 GTGACCGGCGGAGCTTGTGACGCTCACCAGTGGGTGAACATGACAGATGAGGCTCCAG 1020  
Db 1183 GTGACCGGCGGAGCTTGTGACGCTCACCAGTGGGTGAACATGACAGATGAGGCTCCAG 1242  
Qy 1021 CTCATATGCCACACACAGATGAAGACGTGTAAACCTGCTGGCTACAGTTTGGCGCA 1080  
Db 1243 CTCATATGCCACACACAGATGAAGACGTGTAAACCTGCTGGCTACAGTTTGGCGCA 1302

```
OY 1081 GGCACAAATGATGAGATTTGGGCGCTCAATATTACACAGGAACATACGGAAGGGATCCA 1140
DB 1303 GGCACAAATGATGAGATTTGGGCGCTCAATATTACACAGGAACATACGGAAGGGATCCA 1362
OY 1141 TTCTGGGACACTCTTGGGACCACTCTTGGGAAAACTCTGGAAGAGTTGTAGAGTGT 1200
DB 1363 TTCTGGGACACTCTTGGGACCACTCTTGGGAAAACTCTGGAAGAGTTGTAGAGTGT 1422
OY 1201 CAGAAAACCAACCAATCTCGCTTCACAGTGGAGACTGACGATACCAATCTCTTGGCAA 1260
DB 1423 CAGAAAACCAACCAATCTCGCTTCACAGTGGAGACTGACGATACCAATCTCTTGGCAA 1482
OY 1261 CCAGATATTGTTGATGTTGAGATTGTTACCGTTGGGCTCTTGGCCATCTGATCCCT 1320
DB 1483 CCAGATATTGTTGATGTTGAGATTGTTACCGTTGGGCTCTTGGCCATCTGATCCCT 1542
OY 1321 GGGGAATTACAAACCATCTGCGGAGCAAGATTTCTGAGAGCAATTAAGAAATTTGCA 1380
DB 1543 GGGGAATTACAAACCATCTGCGGAGCAAGATTTCTGAGAGCAATTAAGAAATTTGCA 1602
OY 1381 CTTTATGGGATGAGGATATGACCGTTGTTATCGCAGTCTTAAGCAATGTTTATACACAT 1440
DB 1603 CTTTATGGGATGAGGATATGACCGTTGTTATCGCAGTCTTAAGCAATGTTTATACACAT 1662
OY 1441 TACATTTACCAATATGAAAGATATACAGGCTCAGCGGTACAGGAGCAGCATCTACAATCTAT 1500
DB 1663 TACATTTACCAATATGAAAGATATACAGGCTCAGCGGTACAGGAGCAGCATCTACAATCTAT 1722
OY 1501 GGACCAACACACCTGCTCTGCATACATCCAACTCTTGAAGACCTTCTTAAGCAATTTGCT 1560
DB 1723 GGACCAACACACCTGCTCTGCATACATCCAACTCTTGAAGACCTTCTTAAGCAATTTGCT 1782
OY 1561 ACGGACACAGTACCCAACTAGAGAGTGTCCCGAGCTCTCAATCTTCAAAAATCTAATA 1620
DB 1783 ACGGACACAGTACCCAACTAGAGAGTGTCCCGAGCTCTCAATCTTCAAAAATCTAATA 1842
OY 1621 GCTTACTTATTTCTTAATATTCGGGATAGAGCAACCAATTTGGCAACATTTTGGGAGTTC 1680
DB 1843 GCTTACTTATTTCTTAATATTCGGGATAGAGCAACCAATTTGGCAACATTTTGGGAGTTC 1902
OY 1681 TTGCAAGCCAGCAAAACCTGAATATACAGAGTGGGAGAAAGTGTGAAGTTATTGTAGGC 1740
DB 1903 TTGCAAGCCAGCAAAACCTGAATATACAGAGTGGGAGAAAGTGTGAAGTTATTGTAGGC 1962
OY 1741 GCTAAACCAAGAAATTCAGACAGAGAAACGACCCATCAACCTCTCTCACTGTGAGAAA 1800
DB 1963 GCTAAACCAAGAAATTCAGACAGAGAAACGACCCATCAACCTCTCTCACTGTGAGAAA 2022
OY 1801 TACGAGAGACTCTGTAGCTAGCTGCGAGATATGTATACGATGCTCTCGGAGAGCAGAG 1860
DB 2023 TACGAGAGACTCTGTAGCTAGCTGCGAGATATGTATACGATGCTCTCGGAGAGCAGAG 2082
OY 1861 TTTTATTTGGCACAAGAAAGTACTGGGCTGAGCAATGCAACATATACAGGATATTTCA 1920
DB 2083 TTTTATTTGGCACAAGAAAGTACTGGGCTGAGCAATGCAACATATACAGGATATTTCA 2142
OY 1921 GATAGTCCCTACCTGGAATCTACAGAAATAGATATTTTGGACACATCTGGAAGCAGAA 1980
DB 2143 GATAGTCCCTACCTGGAATCTACAGAAATAGATATTTTGGACACATCTGGAAGCAGAA 2202
OY 1981 CTTTGGAAAACCGGCTGTCTACTAGCATTTGAAGCAATTTCTTCTCTTTGAAGTTGTC 2040
DB 2203 CTTTGGAAAACCGGCTGTCTACTAGCATTTGAAGCAATTTCTTCTCTTTGAAGTTGTC 2262
OY 2041 ACTACTTAG 2049
DB 2263 ACTACTTAG 2271
```

RESULT 3  
AAA97632  
ID AAA97632 standard; cDNA; 3108 BP.

```
XX AAA97632:
AC 02-FEB-2001 (first entry)
DT
XX
DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:11.
XX
XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KW cellular ceramide content control; antibody; membrane synthesis;
KW lipid engineering; ceramide metabolism; drug development; ss.
XX
OS Mus sp.
XX
XX WO200058448-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000MO-JP01802.
XX
XX 26-MAR-1999; 99JP-0084743.
XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX Ito M;
XX
XX WPI; 2000-619079/59.
XX
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering
PT reagent for studying structure and functions of ceramide as well as
PT developing drugs for diseases associated with ceramide metabolism -
XX
XX Example 4; Page 55-56; 76pp; Japanese.
XX
XX The invention relates to a mouse neutral/alkaline ceramidase and to
CC nucleic acids encoding it. The invention also relates to expression
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline
CC ceramidase, the recombinant production of the ceramidase, an antibody
CC against the ceramidase, and a method for controlling the ceramide
CC content in cells and/or tissues by introducing the gene or its antisense
CC nucleic acid into the cells and/or tissues. The ceramidase can be used
CC as a lipid engineering reagent for studying structure and functions of
CC ceramide as well as developing drugs for diseases associated with
CC ceramide metabolism. The present sequence represents a clone encoding
CC mouse neutral/alkaline ceramidase.
XX
XX Sequence 3108 BP; 888 A; 760 C; 714 G; 746 T; 0 other;
XX
XX Query Match 100.0%; Score 2049; DB 21; Length 3108;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TTCACTGCTACTACATTTGGCGTGGAGAGCGGATTTGCACAGCAACATGTCAGATATC 60
DB 948 TTCACTGCTACTACATTTGGCGTGGAGAGCGGATTTGCACAGCAACATGTCAGATATC 1007
OY 61 AATTGATGCGCTATGCGCAAAATGCGCAATGCGAGCGGCTCTCTACACAGCGCTTTC 120
DB 1008 AATTGATGCGCTATGCGCAAAATGCGCAATGCGAGCGGCTCTCTACACAGCGCTTTC 1067
OY 121 AGCCGCTTTATATCTGCGGAGATCCAGATGGGTCGAATGCAATGGCATTTTGAAGCGTG 180
DB 1068 AGCCGCTTTATATCTGCGGAGATCCAGATGGGTCGAATGCAATGGCATTTTGAAGCGTG 1127
OY 181 GAACATATGATATTTCCCAAGCACTGAGTGTGAGAGTCTCTGAAGAGACTAAGAGATAA 240
DB 1128 GAACATATGATATTTCCCAAGCACTGAGTGTGAGAGTCTCTGAAGAGACTAAGAGATAA 1187
OY 241 TATGCTCTCTATATGAGAGACAATGTTATCTGAGTGCATTCACACACACTCTGCGC 300
DB 1188 TATGCTCTCTATATGAGAGACAATGTTATCTGAGTGCATTCACACACACTCTGCGC 1247
OY 301 CCAGCAGGTTTTTCCAAATATACACTCTATATACCTGCGCAGGAGGAGATTGCAACCGG 360
```

Db 1248 CCAGCAGGGTTTTCCAAATATACACTATATACCTCCGACGAGGATTCAGCAACCGG 1307  
QY 361 ACCTTTGATACATAGTCTCTGGGATCATGAAGAGCATTTGATATAGCTCACCAATCTT 420  
Db 1308 ACCTTTATGATATAGTCTCTGGGATCATGAAGAGCATTTGATATAGCTCACCAATCTT 1367  
QY 421 AAACGAGGCAAAATCTTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 480  
Db 1368 AAACGAGGCAAAATCTTTATCAACAAAGAAATGTTCTATATGTCAGATCAACCGAAGC 1427  
QY 481 CCCCTCTTACCTCTTGAATCCACAGTCAGAGAGCAGATATTTCTTAAACACAGAC 540  
Db 1428 CCCCTCTTACCTCTTGAATCCACAGTCAGAGAGCAGATATTTCTTAAACACAGAC 1487  
QY 541 AAGGAATGCTGGCTTGAATCTGGAATGTTGAATGAGAGAGCTGGCTTATAGC 600  
Db 1488 AAGGAATGCTGGCTTGAATCTGGAATGTTGAATGAGAGAGCTGGCTTATAGC 1547  
QY 601 TGGTTTGCATCCACCCCGTAGCATGAGCATGAGCAATGCAACCACTTTGATATGATGAT 660  
Db 1548 TGGTTTGCATCCACCCCGTAGCATGAGCATGAGCAATGCAACCACTTTGATATGATGAT 1607  
QY 661 ATGGGCTATGCGGCTTACCTTTTGGAGCAAGAAAGAACAAAGGCTATGCTGAGCAG 720  
Db 1608 ATGGGCTATGCGGCTTACCTTTTGGAGCAAGAAAGAACAAAGGCTATGCTGAGCAG 1667  
QY 721 GGACCGTTTGGAGGCTTGTCTTCATCAAAATCTGGAGAGCGTGTACCCCAATCTT 780  
Db 1668 GGACCGTTTGGAGGCTTGTCTTCATCAAAATCTGGAGAGCGTGTACCCCAATCTT 1727  
QY 781 GGCCCGCATTTGTCAACACAGGAGGAGTCTTGTGACAGCAGCAAGACACCTGTCCCAAC 840  
Db 1728 GGCCCGCATTTGTCAACACAGGAGGAGTCTTGTGACAGCAGCAAGACACCTGTCCCAAC 1787  
QY 841 GGTGGGCTTACATGTCATGATGCCAGCGGACCTGAGACAGATGTTTGGAGCAGCAC 900  
Db 1788 GGTGGGCTTACATGTCATGATGCCAGCGGACCTGAGACAGATGTTTGGAGCAGCAC 1847  
QY 901 ATTATGAGAGGATCATCTATCAAGAGGCAAGAGAGTGTATGCTCTGCTCCAGAG 960  
Db 1848 ATTATGAGAGGATCATCTATCAAGAGGCAAGAGAGTGTATGCTCTGCTCCAGAG 1907  
QY 961 GTGACCGGCCAGTGTCTTGGACGTCAACAGTGGGTGAAACATGACAGATGTGAGGCTCAG 1020  
Db 1908 GTGACCGGCCAGTGTCTTGGACGTCAACAGTGGGTGAAACATGACAGATGTGAGGCTCAG 1967  
QY 1021 CTCATGCCACACACAGTGAAGAGCTGTAAACCTGCTGGGCTAAGATTTGGCGCA 1080  
Db 1968 CTCATGCCACACACAGTGAAGAGCTGTAAACCTGCTGGGCTAAGATTTGGCGCA 2027  
QY 1081 GGCCATTTGATGAGTTCGGGCTCAATATTAACAAGGAATCTAGCAAGAGGATCCA 1140  
Db 2028 GGCCATTTGATGAGTTCGGGCTCAATATTAACAAGGAATCTAGCAAGAGGATCCA 2087  
QY 1141 TTTGGGACACTCTTGGGACACAGCTCTTGGGAAACCATCTGAAGAATTTGTAGAGTGT 1200  
Db 2088 TTTGGGACACTCTTGGGACACAGCTCTTGGGAAACCATCTGAAGAATTTGTAGAGTGT 2147  
QY 1201 CAGAAACCCAAACCAATCTGCTTACAGTGAAGAGCTGACAGATACCAATCTTGGCAA 1260  
Db 2148 CAGAAACCCAAACCAATCTGCTTACAGTGAAGAGCTGACAGATACCAATCTTGGCAA 2207  
QY 1261 CCAGATTTGATGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 1320  
Db 2208 CCAGATTTGATGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 2267  
QY 1321 GGGGATTTAAACAACCATGTCGGGACGAAGATTTTGTGAGGCAATTAAGAAATTTGCA 1380  
Db 2268 GGGGATTTAAACAACCATGTCGGGACGAAGATTTTGTGAGGCAATTAAGAAATTTGCA 2327  
QY 1381 CTTTATGAGTGAAGATATGACCGTTGTTATGCGACAGCTTAAGCAATGTTTATACAT 1440  
Db 2328 CTTTATGAGTGAAGATATGACCGTTGTTATGCGACAGCTTAAGCAATGTTTATACAT 2387

QY 1441 TACATTAACAATATGAAGATATACAGGCTCAGCGGTACGAGGAGCATCTACAAATCTAT 1500  
Db 2388 TACATTAACAATATGAAGATATACAGGCTCAGCGGTACGAGGAGCATCTACAAATCTAT 2447  
QY 1501 GGACCAACACACCTGTCTGATATCATCAACCTCTTTCAGAGACCTTGTGAAGCAATTTGCT 1560  
Db 2448 GGACCAACACACCTGTCTGATATCATCAACCTCTTTCAGAGACCTTGTGAAGCAATTTGCT 2507  
QY 1561 ACGGACACAGTGAAGCAATGAGAGTGTCCGAGGCTCCATTTCTTCAAAAATCTATA 1620  
Db 2508 ACGGACACAGTGAAGCAATGAGAGTGTCCGAGGCTCCATTTCTTCAAAAATCTATA 2567  
QY 1621 GCTTCACTTATTTCTTAATTTTGGGATAGAGCACCATTTGGCAAAATTTTGGGAGTGC 1680  
Db 2568 GCTTCACTTATTTCTTAATTTTGGGATAGAGCACCATTTGGCAAAATTTTGGGAGTGC 2627  
QY 1681 TTGACGACGCAAAACCTGATACAGAGTGGGAGAAAGTGTGAAGTTATATTTGTAGGC 1740  
Db 2628 TTGACGACGCAAAACCTGATACAGAGTGGGAGAAAGTGTGAAGTTATATTTGTAGGC 2687  
QY 1741 GCTAACCCAAAGATTTAGCAGAGAACCCAGACCATTAACCTTCTCACTGTGAGAGAA 1800  
Db 2688 GCTAACCCAAAGATTTAGCAGAGAACCCAGACCATTAACCTTCTCACTGTGAGAGAA 2747  
QY 1801 TAGGAGACCTGTAGCTGACCTGGCAGATATATGTATACATGCTCTGGGAGAGAGG 1860  
Db 2748 TAGGAGACCTGTAGCTGACCTGGCAGATATATGTATACATGCTCTGGGAGAGAGG 2807  
QY 1861 TTTTATTTGGCAAAAGAAATCTGGTCTGAGCAATGCAACAAATATCTGGCATATTTCCA 1920  
Db 2808 TTTTATTTGGCAAAAGAAATCTGGTCTGAGCAATGCAACAAATATCTGGCATATTTCCA 2867  
QY 1921 GATATGCTCTTACCTGGAATCTACAGAAATTAATTTTGGACCAATGCGAAGCAGGAA 1980  
Db 2868 GATATGCTCTTACCTGGAATCTACAGAAATTAATTTTGGACCAATGCGAAGCAGGAA 2927  
QY 1981 CTCTGAAACCCGCTGATACCTGATGATGATTTGAAGAAATTTCTCTTTTGAAGTGTG 2040  
Db 2928 CTCTGAAACCCGCTGATACCTGATGATGATTTGAAGAAATTTCTCTTTTGAAGTGTG 2987  
QY 2041 ACTACTTAG 2049  
Db 2988 ACTACTTAG 2996

RESULT 4  
AAA97635  
ID AAA97635 standard; DNA; 4835 BP.  
XX  
AC AAA97635;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Mouse neutral/alkaline ceramidase clone, SEQ ID NO:16.  
XX  
DE  
XX Neutral/alkaline ceramidase; mouse; murine; recombinant production;  
KW cellular ceramide content control; antibody; membrane synthesis;  
KW lipid engineering; ceramide metabolism; drug development; ds.  
XX  
OS Mus sp.  
OS Synthetic.  
XX  
XX W0200058448-A1.  
XX  
XX 05-OCT-2000.  
XX  
XX 24-MAR-2000; 2000WO-JP01802.  
XX  
XX 26-MAR-1999; 99JP-0084743.  
XX  
XX (TAKI ) TAKARA SHUZO CO LTD.  
XX  
XX







Db 2403 GCTAACCAAGAAATTCAGCAGAGAACAGACCCATCAAACTTCCTCAGCTGTGAGAAA 2462  
QY 1801 TACGAGGACTGTGACTGACTGACGAGATATGATTAACGATGCTCCGAGAGAGAG 1860  
Db 2463 TACGAGGACTGTGACTGACTGACGAGATATGATTAACGATGCTCCGAGAGAGAG 2522  
QY 1861 TTTTATTGGCACAAAGAAATACTGGGTCTGAGCAATGCAACAATATAGTGCATTTCCA 1920  
Db 2523 TTTTATTGGCACAAAGAAATACTGGGTCTGAGCAATGCAACAATATAGTGCATTTCCA 2582  
QY 1921 GATACTGCTTACCTCGGAATCTACAGATATATTTTGGACACAAATCGAAGAGCAA 1980  
Db 2583 GATACTGCTTACCTCGGAATCTACAGATATATTTTGGACACAAATCGAAGAGCAA 2642  
QY 1981 CTTTGAACCCGCTGTACTACTAGCATTTGAAGAAATTTCTCTCCTTTGAAGTTGTC 2040  
Db 2643 CTTTGAACCCGCTGTACTACTAGCATTTGAAGAAATTTCTCTCCTTTGAAGTTGTC 2702  
QY 2041 ACTACTTAG 2049  
Db 2703 ACTACTTAG 2711

RESULT 5  
AAF99977  
ID AAF9977 standard; DNA; 2283 BP.

AC AAF99977;

DT 20-JUL-2001 (first entry)

DE DNA encoding rat protein involved in acid amide bond hydrolysis.

XX Rat; acid amide bond hydrolysis; ceramide; cytosolic; cancer; ds.

OS Rattus norvegicus.

XX Key Location/Qualifiers

FT CDS 1..2283

FT /tag= a

FT /partial

FT /note= "this sequence does not contain a stop codon"

PN JP2001057890-A.

PD 06-MAR-2001.

PF 23-AUG-1999; 99JP-0235218.

PR 23-AUG-1999; 99JP-0235218.

XX (MITU ) MITSUBISHI CHEM CORP.

DR MPI; 2001-304133/32.

DR P-PSDB; AAB97029.

XX Novel protein, used to treat abnormal cell growth e.g. in cancers, is

PT capable of hydrolyzing the acid amide bond between sphingosine and the

PS fatty acid in ceramide.

XX Claim 3: Page 9-11; 14pp; Japanese.

CC The present sequence is provided in a specification relating to a

CC protein having a 761 residue amino acid sequence or its variant

CC comprising a replacement, deletion, insertion, addition or reversion of

CC at least one amino acid, but retaining the activity of hydrolyzing the

CC acid amide bond between sphingosine and the fatty acid in a ceramide.

CC The protein can be used to treat diseases caused by abnormal cell

CC growth such as cancers.

XX Sequence 2283 BP; 641 A; 584 C; 524 G; 534 T; 0 other;

XX Query Match 89.3%; Score 1830; DB 22; Length 2283;

Best Local Similarity 93.4%; Pred. No. 0;  
Matches 1911; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 1 TTCAGTGGCTACTACATTTGGCTTTGGAGAGCGGATTTGCACAGGCAAGTGTCAATATC 60  
Db 238 TTCAGTGGCTACTACATTTGGCTTTGGAGAGCGGATTTGCACAGGCAAGTGTCAATATC 297  
QY 61 AATTGATGGGCTATGGCAAAATGGCCGAATGACAGGGGGCTCTCCACAGGCTGTTC 120  
Db 298 AATTGATGGGCTATGGCAAAATGGCCGAATGACAGGGGGCTCTCTTACACAGGCTGTTC 357  
QY 121 ACCCGCTTTTATCTTGGCGGATCCAGATGGGTCAAAATGCAATGCAATTTGTGACGTG 180  
Db 358 ACCCGCGCTTTATCTTGGCGGATCCAGATGGGTCAAAATGCAATGCAATTTGTGACGTG 417  
QY 181 GAACATATGATATTTTCCACAGCTGAGGTTGGAGGCTCTGAAAGAGCTAGAGGTAA 240  
Db 418 GAACATATGATATTTTCCACAGCTGAGGTTGGAGGCTCTGAAAGAGCTAGAGGTAA 477  
QY 241 TATGGCTCTGTATGCAAGAGACAAATGTTATCTGAGTGCATTCACACACTCTGAC 300  
Db 478 TATGGCTCTGTATGCAAGAGACAAATGTTATCTGAGTGCATTCACACACTCTGAC 537  
QY 301 CCAGCAGGGTTTTTCCAAATATACACTATATATCTGCGCAGGAGGATTCAGCAACCG 360  
Db 538 CCAGCAGGATTTTCCAAATATACACTATATATCTGCGCAGGAGGATTCAGCAACCG 597  
QY 361 ACCCTTACATATGCTCTGAGATCATGAAGACATTTATATGCTCACAAATCTT 420  
Db 598 ACCCTTACATATGCTCTGAGATCATGAAGACATTTATATGCTCACAAATCTT 657  
QY 421 AATCCAGCAAAATGTTTATCAACAAGAAATGTTGTAATGTCAGATCAACGGAAG 480  
Db 658 AATCCAGCAAAATGTTTATCAACAAGAAATGTTGTAATGTCAGATCAACGGAAG 717  
QY 481 CCTCCTCTTACCTTCTGATTCACAGTACAGAGAGCAAGGATTTCTTCAACACAGAC 540  
Db 718 CCTCCTCTTACCTTCTGATTCACAGTACAGAGAGCAAGGATTTCTTCCACACAGAC 777  
QY 541 AAGCAATGCTGCTTGAAGTGTGATTTGAATGAGAAAGACTTGGGCTTTATACG 600  
Db 778 AAGCAATGCTGCTTGAAGTGTGATTTGAATGAGAAAGACTTGGGCTTTATACG 837  
QY 601 TGGTTGCCATCCACCCGCTGAGTGAACATATCAACCACTTTGTGTAATGTATACAT 660  
Db 838 TGGTTGCCATCCACCCGCTGAGTGAACATATCAACCACTTTGTGTAATGTATACAT 897  
QY 661 ATGGCTATGCGGCTTACCTTTTGGAGAAAGAAAGCAAGGCTATCTGCTGAGAGC 720  
Db 898 ATGGCTATGCGGCTTACCTTTTGGAGAAAGAAAGCAAGGCTATCTGCTGAGAGC 957  
QY 721 GGACCGTTTGTAGCAGGCTTTGCTTATCAATTCGAGAGCTGTACCCAACTTCTT 780  
Db 958 GGACCGTTTGTAGCAGGCTTTGCTTATCAATTCGAGAGCTGTGTCCCAACTTCTT 1017  
QY 781 GGCCCGCAATGTTGTCAACACAGGGAGTCTTGTCAACAGCAAGAGCAACCGTCCAC 840  
Db 1018 GGCCCGCAATGTTGTCAACACAGGGAGTCTTGTGTCAACAGCAAGAGCAACCGTCCAC 1077  
QY 841 GGTGGGCTTACGATGTGATGCGCAGCGGACTGTGACAAAGACATGTTTGTAGAGCACAC 900  
Db 1078 GGTGGGCTTACGATGTGATGCGCAGCGGACTGTGACAAAGATGTTTGTAGAGCACAC 1137  
QY 901 ATTATAGACGATATCTATATGAAGGCAAGAGGCTGTATGCTGTGCTCCAGAG 960  
Db 1138 ATTATAGACGATATCTATATGAAGGCAAGAGGCTGTATGCTGTGCTCCAGAG 1197  
QY 961 GTGACCGGCGGCTGCTTGCACCTCACAGGATGGTGAACATATGATGTGAGCTCCAG 1020  
Db 1198 GTGACCGGCGGCTGCTTGCACCTCACAGGATGGTGAACATATGATGTGAGCTCCAG 1257  
QY 1021 CTCATGCGCACACACAGTGAAGCTGTAAACCTGCTGCGCTACAGTTTGTGCGCA 1080  
Db 1257 CTCATGCGCACACACAGTGAAGCTGTAAACCTGCTGCGCTACAGTTTGTGCGCA 1080

Db 1258 CTCATGCGACACACAGAGTGAAGAGCTGTAAGCCGCCCTGGGCTACAGTTTGGCCGCA 1317  
Qy 1081 GGCACAATTGATGAGATTTGGGCGCTCAATATTACAGAGGACATACGGAAGGGGATCCA 1140  
Db 1318 GGCACAATTGATGAGATTTGGGCGCTCAATATTACAGAGGACATACGGAAGGGGATCCA 1377  
Qy 1141 TTTCGGGACACTCTTCGGGACAGAGCTCTTGGGAAAACACTCTGAAGATTTGTAGAGTGT 1200  
Db 1378 TTTTGGGACACTCTTCGGGACAGAGCTCTTGGGAAAACACTCTGAAGATTTGTAGAGTGT 1437  
Qy 1201 CAGAAACCCAAACCAATCTGCTTCACAGTGAAGAGCTGACAGATGACATCTTGGGCAA 1260  
Db 1438 CAGAAACCTTAACCAATCTGCTTCACAGTGAAGAGCTGACAGATGACATCTTGGGCAA 1497  
Qy 1261 CCAGATATTTGTGATGTTGATGATTTACCGTTGGGCTCTTGGCCATAGCTGATTCCT 1320  
Db 1498 CCAGATATTTGTGATGATTTGATGATTTACCGTTGGGCTCTTGGCCATAGCTGATTCCT 1557  
Qy 1321 GGGGAATTAACCAATCTGCTTCAGAGGACGAAATTTGTCAGGCAATTAAGAAATTTGCA 1380  
Db 1558 GGGGAATTTACCAATCTGCTTCAGAGGACGAAATTTGTCAGGCAATTAAGAAATTTGCA 1617  
Qy 1381 CTTTATGGATGAAGATATGACGCTGTTATGCGAGCTCTTAAGCAATGTTTATACAT 1440  
Db 1618 CTTTATGGATGAAGATATGACGCTGTTATGCGAGGCTGACCAATGTTTATACCAT 1677  
Qy 1441 TACATTACCAATATGAAGAATACAGAGCTCAGCGGTACGAGCAGCATCTACATCTAT 1500  
Db 1678 TACATTACCAATATGAAGAATACAGAGCTCAGCGGTATGAGAGCATCTACATCTAT 1737  
Qy 1501 GGACCAACACCCCTGCTGCTGATCATCTCAACTCTTCAGAGACCTTGTGAAGCAATTTGCT 1560  
Db 1738 GGACCAACACCCCTGCTGCTGATCATCTCAACTCTTCAGAGACCTTGTGAAGCAATTTGCT 1797  
Qy 1561 ACGGACACATAGCAATGACAGAGTGTGCTCCGAGCTCCTCAATCTTCAAAAATCTATATA 1620  
Db 1798 ACGGATACATAGCAATGACAGAGTGTGCTCCGAGCTCCTCAATCTTCAAAAATCTATATA 1857  
Qy 1621 GCTTCACTTATTCCTTAATATTTGGGATAGAGACCAATTTGGCAAACTTTTGGGATGTC 1680  
Db 1858 GCTTCACTTATTCCTTAATATTTGGGATAGAGACCAATTTGGGATGTC 1917  
Qy 1681 TTGCGAGCGCAAAAACCTGAAATACAGAGTGGGAGAAAGTGGTGAAGTTATTTGTAGGC 1740  
Db 1918 TTGCGAGCGCAAAAACCTGAAATACAGAGTGGGAGAAAGTGGTGAAGTTGTGTTGAGGC 1977  
Qy 1741 GCTAACCCAAAGAAATTCAGCAGAGAACCAAGCCATCAAACTTCTCAGCTGTGAGAGAA 1800  
Db 1978 GCTAACCCAAAGAAATTCAGCAGAGAACCAAGCCATCAAACTTCTCAGCTGTGAGAGAA 2037  
Qy 1801 TACGAGAGCTCTGTAGCTGACGCGACAGATATATGATACAGATGCGCTCTGGAGACGAGG 1860  
Db 2038 TACGAGAGCTCTGTAGCGCAACCTGAGATATGATCAATGATCGCTCTGGAGACGAGG 2097  
Qy 1861 TTTTATTTGGCAAAAGGAAATCTGGGCTGAGCAATGCAACATATTTGGCATATTTCCA 1920  
Db 2098 TTTTATTTGGCAAAAGGAAATCTGGGCTGAGCAATGCAACATATTTGGCATATTTCCA 2157  
Qy 1921 GATACGCTGCTACCCCTGGAATCTACAGAAATATATTTGGACACATCGGAAGCAGAA 1980  
Db 2158 GATACGCTGCTGCTGAGCTGAGCTACAGAAATATTTGGACACATCGGAAGCAGAA 2217  
Qy 1981 CTTTGTGAACCCGCTGTCTATAGCATATTTGAAGAAATTTCTTCTCTTTGAAGTTGTC 2040  
Db 2218 CTTTGAAGCCTGCTGTCTATAGCATATTTGAAGAAATTTCTTCTCTTTGAAGTTGTC 2277  
Qy 2041 ACTACT 2046  
Db 2278 ACTACT 2283

ID AAD22668 standard; cDNA: 2507 BP.  
XX  
AC AAD22668;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE . Human ceramidase cDNA.  
XX  
XX Human; ceramidase; nephrotropic; antipsoriatic; antisense therapy;  
KW gene therapy; proliferative disorder; cancer; cardiovascular disease;  
KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;  
KW notropic; signal transduction; breast cancer; autoimmune disorder;  
KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;  
KW glomerular disease; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 39..2324  
FT /tag= a  
FT /product= "Human ceramidase protein"  
FT /note= "There is an additional stop codon from position  
FT 2325-2327"  
FT sig\_peptide 39..95  
FT /tag= b  
FT mat\_peptide 96..2321  
FT /tag= c  
FT /product= "Mature ceramidase protein"  
PN M0200155410-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 29-JAN-2001; 2001MO-US02866.  
XX  
PR 28-JAN-2000; 2000US-178975P.  
XX  
PA (MUSC-) MUSC FOUND RES DEV.  
XX  
PI Hannun YA, El Bawab S;  
XX  
PI WPI: 2002-025687/03.  
XX  
DR P-PSDB: AAEI3544.  
XX  
XX Human mitochondrial ceramidase protein and gene, modulation of which is  
XX useful for preventing and treating proliferative disorders e.g. cancer,  
XX cardiovascular disease, inflammation and neurodegenerative disorders -  
XX  
XX Claim 1; Fig 3; 113pp; English.  
XX  
XX The present sequence is a cDNA encoding human mitochondrial ceramidase  
XX protein. Ceramidase protein and gene are useful for treating a disease  
XX or disorder involving cell over proliferation or sphingolipid signal  
XX transduction especially breast cancer, cardiovascular disorder or  
XX inflammation. Ceramidase protein and gene are useful for treating  
XX disorders involving deficient cell proliferation or growth e.g.  
XX neurodegenerative disorders (Alzheimer's disease), growth deficiencies  
XX and lesions. Ceramidase protein is also useful for diagnosis of  
XX hyperproliferative diseases. Ceramidase gene can be used as an immunogen  
XX to generate antibodies which are useful for diagnosis, prevention and  
XX treatment of hyperproliferative diseases and for detecting ceramidase  
XX gene product in a biological sample. The hyperproliferative disorders  
XX include cancers and autoimmune disorders such as lupus nephritis,  
XX glomerular disease. Ceramidase gene is also useful in antisense therapy  
XX and gene therapy.  
SQ Sequence 2507 BP; 720 A; 583 C; 567 G; 637 T; 0 other;  
Query Match 73.8%; Score 1511.4; DB 24; Length 2507;  
Best Local Similarity 83.6%; Pred. No. 0;  
Matches 1713; Conservative 0; Mismatches 336; Indels 0; Gaps 0;  
Qy 1 TTTCAGTGCTACTACATTTGGCGTGGAGAGCGGATTTGCACAGGACAGTGTACATATTC 60





Db 45 AACAGGCTCAGCATATGAGCAGCATGACAAATTTATGAGCCGACCCATTATCTGCTT 104  
 Oy 1523 ACATTCACATCTTGAAGACCTTGTGAAGCAATTTGACAGACAGATAGCAACATCA 1582  
 Db 105 ACATTCAGCTTTGAGAACTTGTGATGAGTATGAGTACAGACAGGATAGCAACCTGA 164  
 Oy 1583 GCAGTGTGCTCCAGGCTCCATCTTCCAAATCTAATAGTCTCACTTTCTCTAATATG 1642  
 Db 165 GCAGAGGTCCAGAACTCCCTTTTCAACAATTAATAGTCTCACTTTCTCTAATATG 224  
 Oy 1643 CGATAGACGACCAATTTGGCAAACTTTGGGATGCTTCCAGCAGCAAAACCTGAAT 1702  
 Db 225 TGAGTAGAGCACCAAAAGGAGAACTTTGGGAGATGCTCCAGCAGCAAAACCTGAAT 284  
 Oy 1703 ACAGATGGGAGAGTGTGATATTTGATAGGCGCAACCAAGATTCAGAG 1762  
 Db 285 ACAGATGGGAGAGTGTGATATTTGATAGGCGCAACCAAGATTCAGAG 344  
 Oy 1763 AGAACGAGACCCATCAAACTTCTCACTGTGAGAAATACGAGACTTGTAGCTGACT 1822  
 Db 345 AAAACGAGACCCATCAAACTTCTCACTGTGAGAAATACGAGACTTGTAGCTGACT 404  
 Oy 1823 GCGAGATATATATACGATCTCTCTGGAGAGACAGGTTTATTTGGCAAAAGATATAC 1882  
 Db 405 GCGAGATATATATACGATCTCTCTGGAGAGACTTCTTTATTTGGCAAAAGGACTCC 464  
 Oy 1883 TGGGCTGAGCAATGCAACATATACGCAATATTCGCAATATCCGCTTGAATCT 1942  
 Db 465 TGGGCTGAGCAATGCAACATATTCGCAATATTCGCAATATCCGCTTGAATCT 524  
 Oy 1943 ACAGATATATATATTTGAGCAACATGCAAGCAAGCACTTCTGAAACCCGCTGATAC 2002  
 Db 525 ACAGATATATATATTTGAGCAACATGCAAGCAAGCACTTCTGAAACCCGCTGATAC 584  
 Oy 2003 TAGCATTTGAAGAAATTTCTCTCTTTGAAGTTGCTACTACTTAC 2049  
 Db 585 TTTTATTGGAAGCACTTCCCGGCTTTTGAAGTTGTAATATTTAG 631  
 RESULT 9  
 AAZ06239 standard; DNA: 1194 BP.  
 ID AAZ06239;  
 AC AAZ06239;  
 DT 30-SEP-1999 (first entry)  
 DE Human secreted protein gene No. 21.  
 XX Human: secreted protein; fusion protein; gene therapy; protein therapy;  
 KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KM immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;  
 KM inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 XX  
 PN WO9335158-A1.  
 PD 15-JUL-1999.  
 XX  
 PF 06-JAN-1999; 99WO-US00108.  
 XX  
 PR 07-JAN-1998; 98US-0070704.  
 PR 07-JAN-1998; 98US-0070657.  
 PR 07-JAN-1998; 98US-0070658.  
 PR 07-JAN-1998; 98US-0070692.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX  
 PI Brewer LA, Duan RD, Edner R, Lafleur DW, NI J;  
 PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;  
 DR WPI: 1999-444190/37.  
 DR P-PSDB: AAY38406, AAY38464, AAY38465, AAY38466, AAY38467, AAY38468,  
 DR AAY38469, AAY38470, AAY38471.  
 PS New isolated human genes and the secreted polypeptides they encode  
 PS Claim 1: Page 160; 227pp: English.  
 CC This sequence represents a nucleic acid molecule which encodes a  
 CC secreted human protein. The gene number is given in the descriptor line.  
 CC The gene can be used to generate fusion proteins by linking to the gene  
 CC to a human immunoglobulin Fc portion (e.g. AAZ06210) for increasing the  
 CC stability of the fused protein as compared to the human protein (nucleic  
 CC acid sequences: AAZ06219-206263; amino acid sequences AAY38386-Y38498)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological  
 CC conditions can be diagnosed by determining the amount of the new  
 CC polypeptides in a sample or by determining the presence of mutations in  
 CC the new polynucleotides. Specific uses are described for each of the 36  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAZ06219 for described uses).  
 SQ Sequence 1194 BP; 358 A; 291 C; 249 G; 296 T; 0 other:  
 Query Match 21.1%; Score 431.8; DB 20; Length 1194;  
 Best Local Similarity 81.1%; Pred. No. 1.8e-125;  
 Matches 502; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
 Oy 1 TTCAGTGGCTACTACATTTGGCGTGGAGAGCGGATTTGCACAGACAGTGTGATATC 60  
 Db 426 TTCAGTGGCTACTACATTTGGCGTGGAGAGCGGATTTGCACAGACAGTGTGATATC 485  
 Oy 61 AATTGATGGGCTATGCAAAATATGCGCAATGACGAGGCTCTCCACAGGCTGTTC 120  
 Db 486 AATTGATGGGCTATGCAAAATATGCGCAATGACGAGGCTCTCCACAGGCTGTATAC 545  
 Oy 121 AGCGTGGCTTATCTTGGCGGATTCAGATGGGCTCAATGCAATGGCATTTGTGAGCGTG 180  
 Db 546 AGTGGTGGCTTATCTTGGCGGATTCAGATGGGCTCAATGCAATGGCATTTGTGAGCGATC 605  
 Oy 181 GAATATATATATTTCCCAACGACTGAGTGTGAGGCTCTGGAAGACTAGAGCTTAA 240  
 Db 606 GACATAGGCAATGATATCCCAAGGCTGAGGCTGAGGCTCTGGAAGACTAGAGCTTAA 665  
 Oy 241 TATGCTCTCTGTATGCAAGAGACAAATTTATCCGATGCGCATTCACACACACTCTGCG 300  
 Db 666 TATGCTCTCTGTATGCAAGAGAGATATATATATATATATATATATATATATATATAT 725  
 Oy 301 CCAGAGGAGTTTTCAT 360  
 Db 726 COTGAGAGAT 785  
 Oy 361 ACCTTTCAT 420  
 Db 786 ACTTTCAT 845  
 Oy 421 AAACGAGCAAAATTTAT 480  
 Db 846 AAACGAGCAAAATTTAT 905  
 Oy 481 CCTCTCTTACCTTCTGAATTCACAGTACAGAGAGCAAGGATTTCTTCAACACAGAC 540  
 Db 906 CCGTATCTTACCTTCAAAATTCGAGTACAGAGAGCAAGGATTTCTTCAAAATACAGAC 965  
 Oy 541 AAGGAATGCTGCTTGAATGCTGATTTGAATGAGAGAGACTTGGGCTTATATACG 600  
 Db 966 AAGGAATGATATGTTTGAATGATATTTGAATGAGAGAGACTTGGGCTTATATACG 1025

QY 601 TGGTTGGCATCCACCCCG 619  
DB 1026 TTTTCATTCAGCAAGTCTG 1044

RESULT 10  
ID AAF99979 standard; cDNA; 539 BP.  
XX AAF99979;  
XX AAF99979;  
XX 20-JUL-2001 (first entry)  
DE PCR product isolated from rat cDNA.  
XX  
XX Rat; acid amide bond hydrolysis; ceramide; cytosolic; cancer; ss.  
XX  
XX Rattus norvegicus.  
XX JP2001057890-A.  
XX  
XX 06-MAR-2001.  
XX  
XX 23-AUG-1999; 99JP-0235218.  
XX  
XX 23-AUG-1999; 99JP-0235218.  
XX  
XX (MITU) MITSUBISHI CHEM CORP.  
XX  
XX WPI; 2001-304133/32.  
XX  
XX Novel protein, used to treat abnormal cell growth e.g. in cancers, is  
PT capable of hydrolyzing the acid amide bond between sphingosine and the  
PT fatty acid in ceramide -  
XX  
XX Example 3; Page 14; 14pp; Japanese.  
XX  
XX The present sequence is provided in a specification relating to a  
CC protein having a 761 residue amino acid sequence or its variant  
CC comprising a replacement, deletion, insertion, addition or reversion of  
CC at least one amino acid, but retaining the activity of hydrolyzing the  
CC acid amide bond between sphingosine and the fatty acid in a ceramide.  
CC The protein can be used to treat diseases caused by abnormal cell  
CC growth such as cancers.  
XX  
XX Sequence 539 BP; 181 A; 110 C; 113 G; 135 T; 0 other;

Query Match 19.3%; Score 396.4; DB 22; Length 539;  
Best Local Similarity 94.1%; Pred. No. 1.8e-114;  
Matches 412; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1612 AATCTAATAGCTTCACTTATTTCTTAATATTTGGGATGAGCACCATTGGCAAAATTTT 1671  
DB 1 AACTGTATAGTTCACCTATCTATCTTAATTTGCGATGAGCACCATTAGCAAAATTTT 60  
QY 1672 GGGGATGTCTTGGAGCCAGCAAAACCTGAATACAGACTGGAGAAAGTGTGAAGTTATA 1731  
DB 61 GGGGATGTCTTGGAGCCAGCAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATC 120  
QY 1732 TTTGTAGGCGCTTAACCAAGAATTCAGCAGAGAACCAACCATCAACTTCTCTACT 1791  
DB 121 TTTGTAGGCGCTTAACCAAGAATTCAGCAGAGAACCAACCATCAACTTCTCTACT 180  
QY 1792 GTGAGAAATATCAGAGACTCTGTAGCTGAGTATGTATATGATGCTGCTCTG 1851  
DB 181 GTGAGAAATATCAGAGACTCTGTAGCTGAGTATGTATATGATGCTGCTCTCTG 240  
QY 1852 GAGACAGAGTTTATTTGGCAAAAGAAATCTGGTCTGAGCAATGCAACATATATCTGG 1911  
DB 241 GAGACAGAGTTTATTTGGCAAAAGAAATCTGGTCTGAGCAATGCAACATATATCTGG 300  
QY 1912 CATATTCAGATATCTGCTTACCTCGAATCTACAGAAATTAAGATATTTTGGACCAATCGG 1971

DB 301 CATATTCAGATATCTGCCCTCCCTGGAGTCTACAGATCATGATATTTTGGACAAATCGG 360  
QY 1972 AAGCAGAACTTCTGAAACCCGCTGTCATCTAGCATTTTGAAGAAATTTCTTCCCTTT 2031  
DB 361 AAGCAGAACTTCTGAAACCCGCTGTCATCTAGCATTTTGAAGAAATTTCTTCCCTTT 420  
QY 2032 GAAGTGTCTCACTACTTAG 2049  
DB 421 GAATTTCTCACTACTTAG 438

RESULT 11  
ID ABL02671 standard; cDNA; 2553 BP.  
XX ABL02671;  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 2495.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
XX WPI: 2001-656860/75.  
XX P-PDB: ABB58568.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 2495; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2553 BP; 659 A; 685 C; 679 G; 530 T; 0 other;

Query Match 16.3%; Score 334.8; DB 23; Length 2553;  
Best Local Similarity 53.4%; Pred. No. 1.4e-94;  
Matches 842; Conservative 0; Mismatches 697; Indels 39; Gaps 5;

QY 9 CTACTACATTTGGCGTTGGAGAGCGGATTCACAGACAGTCTCAATATCAATTTGAT 68  
DB 510 CTACAAAGTGGCGTGGCGCCGCGGATATCACAGACCTCCAGTGGAGATCAATTTGAT 569  
QY 69 GGGCTATGGCAAAATATGCGCAGATGACAGCGGCTCTCCACAGAGCTGTCCAGCGGTGC 128  
DB 570 GGGCTATGGCAATCAAGCAAGTGGTCTGTCATCACACCCGTTGCTTGGCCGCTGC 629











PN EP980912-A1.  
 XX 23-FEB-2000.  
 XX 20-AUG-1999; 99EP-0116154.  
 XX 20-AUG-1998; 98EP-0234769.  
 PA (TAKI ) TAKARA SHUZO CO LTD.  
 PI Okino N, Ito M;  
 XX WPI: 2000-258590/23.  
 DR P-PSDB: AAT57314.  
 XX  
 PT New Pseudomonas aeruginosa alkaline ceramidase gene, useful for  
 PT diagnosis, treatment and prevention of atopic dermatitis -  
 XX  
 PS Example 1; Page 27-28; 32pp; English.  
 XX  
 CC This DNA encodes a ceramidase from Pseudomonas aeruginosa strain AN-17.  
 CC Ceramidase can be produced by standard recombinant methodology.  
 CC Ceramidase converts ceramide to sphingosine and a fatty acid, causes, or  
 CC exacerbates, atopic dermatitis. The ceramidase gene, polypeptide, derived  
 CC oligonucleotides and ceramidase-specific antibodies are used to prevent  
 CC and/or treat atopic dermatitis. The gene and the polypeptide can be used  
 CC as reagents for lipid engineering and for analyzing the structure and  
 CC activity of ceramides.  
 XX  
 SO Sequence 2013 BP; 393 A; 700 C; 591 G; 329 T; 0 other;

Query Match 5.9%; Score 121.2; DB 21; Length 2013;  
 Best Local Similarity 49.7%; Pred. No. 4.7e-27;  
 Matches 340; Conservative 0; Mismatches 338; Indels 6; Gaps 1;

OY 9 CTACTACATGTCGGCTGGAGAGCGGATTCGACAGACAGTGTGATATCAATTTGAT 68  
 DB 84 CTACCGCTTCGGCTGGAGAGCGGATTCGACAGACAGTGTGATATCAATTTGAT 143  
 OY 69 GGGCTATGCGCAAAATGTCGACAGATGCGGCTCTCCACAGAGCTGTCAGCCGTC 128  
 DB 144 GGGTACTGCTCCCTCGAAGACAGACAGCGCGCATCCACATGCCAGTGGCGCTGC 203  
 OY 129 TTTTATCTTGGCGGATCCAGATGCGTCAATCGAATGCGATTTGAGCGTGAATATG 188  
 DB 204 CTTCCTGATCGAGGAGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 263  
 OY 189 TATGATTTCCCAAGACAGTGTGAGAGTCTGAAGAGAGTAAATATATGCTC 248  
 DB 264 GATGACCTTCAGGCGGCTGACCTGAAGTCTGCGCCGCTCAAGCGCAAGTACCCGG 323  
 OY 249 TCTGTATGAGAGACAAATGTTATCTGAGTGCATTCACACACTCTGCGCCAGCAG 308  
 DB 324 TGTACGACGAGAACAGTATGTCGCGCCGACACCCACTCCGCTCGCGGCGG 383  
 OY 309 GTTTTCCATATACATCTATATCTGCGCCAGCGAGGATTCAGCAACCGACCTTCA 368  
 DB 384 CTTCCTCCACTAGCGGATGATCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 443  
 OY 369 GTACATAGTCTGAGATCATGATGATATGATGATGATGATGATGATGATGATGATG 428  
 DB 444 CGCCATCTGTCAGCGGATGCTCGCTCAATCGAGCGGCGCCAGGTTGACAGCCGG 503  
 OY 429 CAAATCTTTATCAAAAGAAATGTTGCTAATGTGCAATCAACCGAAGCCCTCTC 488  
 DB 504 CGGCTGTTCTAGCGGAGGAGCTGGGCAAGCGCAACCGCAACGTTGCTGCTGTC 563  
 OY 489 TTACTTCTGAATCCAGATCAGAGAGCAAGTATTTCTTAACACAGACAGAAAT 548  
 DB 564 GCACCTGAAGATC-----CGACATGCGCGGCTACGAGATGCAATGCAACCGCAGAT 617  
 OY 549 GCTGCTTGAAGAGTGTGATTTGAATGAGAAAGACTTGGTCTTATCAGCTGCTTGC 608

DB 618 GAGCGTCTACCTTCGTGAGAGCCAGCGGAGCTGGCGGCGGATCAGTTGTTCC 677  
 OY 609 CATCCACCCCGTAGAGCATAGACAACTTTGTGAATAGTACATATATGGGCTA 668  
 DB 678 GGTGACAGCAGCTGATGATGACCAAGCCCAATCAGCTGATCTCCCGGACAAAGGCTA 737  
 OY 669 TCGGCTTACCTTTTGGAGCAGA 692  
 DB 738 CGCTCTTATCACTGGAGACAGCA 761

Search completed: July 5, 2003, 09:08:47  
 Job time : 474 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 5, 2003, 08:54:54 ; Search time 5393 Seconds

(without alignments)  
11057.231 Million cell updates/sec

Title: US-09-937-521-15

Perfect score: 2049

Sequence: 1 ttcaagtgctactactgg.....ttgaagtgctactactg 2049

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: GenEmbl.\*  
2: gb\_ba.\*  
3: gb\_htg.\*  
4: gb\_in.\*  
5: gb\_com.\*  
6: gb\_ov.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rnd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	100.0	3108	10	AB037111	AB037111 Mus muscu
2	2047.4	99.9	4835	10	AB037181	AB037181 Mus muscu
3	1833	89.5	2456	10	AB057433	AB057433 Rattus no
4	1830	89.3	2283	6	E50433	E50433 Novel prote
5	1509.8	73.7	2507	6	AX207121	AX207121 Sequence
6	1508.2	73.6	2289	9	AF250847	AF250847 Homo sapi
7	1493	72.9	1571	10	BC022604	BC022604 Mus muscu
8	396.4	19.3	539	6	E50435	E50435 Novel prote
9	264.6	12.9	2140	3	DD082513	DD082513 Dictyosteli
10	192.4	9.4	165407	2	AC117816	AC117816 Mus muscu
11	179.8	8.8	165407	2	AC117816	AC117816 Mus muscu
12	179.8	8.8	184101	2	AC113485	AC113485 Mus muscu
13	158	7.7	235	9	HUMDAC12M5	D17062 Human HepG2
14	155.2	7.6	53059	2	AC012788	AC012788 Drosophi
15	155.2	7.6	173844	3	AC008299	AC008299 Drosophi
16	155.2	7.6	238986	3	AE003774	AE003774 Drosophi
17	150.4	7.3	162228	9	AL450382	AL450382 Human DNA
18	150.4	7.3	176865	2	AL513306	AL513306 Homo sapi
19	150.4	7.3	176899	9	AL589794	AL589794 Human DNA
20	143.6	7.0	2618	8	AY057506	AY057506 Arabidops
21	135.2	6.6	105838	2	AC105714	AC105714 Rattus no
22	122.8	6.0	2013	1	PAE315932	PAE315932 Pseudomon
23	122.8	6.0	11622	1	AE004519	AE004519 Pseudomon
24	122.2	6.0	110000	2	AC055726_1	Continuation (2 of
25	122.2	6.0	175330	2	AC012131	Continuation (2 of
26	122.2	6.0	187130	2	AL583858	AL583858 Homo sapi
27	121.2	5.9	1941	6	AR162242	AR162242 Sequence
28	121.2	5.9	1941	6	E62982	E62982 Ceramidase
29	121.2	5.9	2013	6	AR162249	AR162249 Sequence
30	121.2	5.9	2013	6	E62989	E62989 Ceramidase
31	121.2	5.9	2744	1	AB028646	AB028646 Pseudomon
32	115.2	5.6	110000	2	AC055726_2	Continuation (3 of
33	105.6	5.2	77758	8	AB016885	AB016885 Arabidops
34	105.4	5.1	110000	2	AC055726_0	Continuation (2 of
35	105.4	5.1	110000	2	AC055726_1	Continuation (2 of
36	100.6	4.9	175330	2	AC012131	Continuation (2 of
37	80	3.9	19770	1	MRC1376	Z95972 Mycobacteri
38	77.6	3.8	565	6	AX094283	AX094283 Sequence
39	75.6	3.7	153794	8	NC12344	AL67009 Neurospor
40	74.2	3.6	19352	1	AE006964	AE006964 Mycobacte
41	72.4	3.5	184101	2	AC113485	AC113485 Mus muscu
42	72	3.5	154561	8	AP002819	AP002819 Oryza sat
43	71	3.5	200	11	DM208105	Z50514 D. melanoga
44	70.2	3.4	104830	8	AC022464	AC022464 Genomic s
45	67.6	3.3	105838	2	AC105714	AC105714 Rattus no

## ALIGNMENTS

RESULT 1  
AB037111 3108 bp mRNA linear ROD 08-JUN-2001  
LOCUS AB037111  
DEFINITION Mus musculus LCBASE mRNA for neutral ceramidase, complete cds.  
ACCESSION AB037111  
VERSION AB037111.1 GI:7576308  
KEYWORDS neutral ceramidase.  
SOURCE Mus musculus liver cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (sites)  
Tanil,M., Okino,N., Mori,K., Tanigawa,T., Izu,H. and Ito,M.  
Molecular cloning of the full-length cDNA encoding mouse neutral  
ceramidase. A novel but highly conserved gene family of

neutral/alkaline ceramidases  
J. Biol. Chem. 275 (15), 11229-11234 (2000)  
MEDLINE 20219171  
REFERENCE 2 (bases 1 to 3108)  
AUTHORS Okino, N. and Ito, M.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2000) Makoto Ito, Kyushu University, Graduate  
School of Bioresource and Biomedical Sciences, Department of  
Bioscience and Biotechnology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka  
812-8581, Japan (E-mail:makotoi@agr.kyushu-u.ac.jp,  
Tel:81-92-642-2900, Fax:81-92-642-2900)

## FEATURES

source

1..3108  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="liver"  
1..3108  
/gene="Lcdase"  
726..2996  
/gene="Lcdase"  
/codon\_start=1  
/product="neutral ceramidase"  
/protein\_id="BAA94545.1"  
/db\_xref="GI:7576309"

## CDS

/translation="MAKRFTLEAFLEFLVIMYTVALLTLEFVTSCTENKDS  
GNHWFSTLGGSTTPTPTTPNPFSEFNFSGYIGVRADCTQVSDINLMGYN  
GOMAGLRLRLFSRAFLADPDGSNMAVSVELMIQLRLLEYKRLSEYKSLR  
RDNLVLSAHTHSAGPAGFOYLYLILASGEFNRRTQYVSGIMKSIDAHNLKPK  
IFINKGNVANOINRSPSYLLNPOSERYSNIDKEKLVLDLNGEDLGLSMF  
AHPVSMNSNHFVNDMGVAAFLPEQKNKGYLPGCGPFAVASSLGDVSPNII  
GPHCVNTESSCNDKSTCNGSPSMCMAAGPCQDMFESHIIIGRIIYOKAKELYSAS  
QEVTVLAHQMVMNTDYSVOLNATHVTKCPALDGSFAAGTIDVSGMLTGGIT  
EGDPMDVLRDLGKPSSEIVECQPKPILHSELTPHMOPIYDVOYVWGL  
ALAAIPGELTMSGRFRFRAIKKEFALYGMKDMYVIALSIPVHYHLYTTEYOQR  
YEAATYIGPHTLSAVIOLFRLAKAIADTVANNSGSEPEPFHLLSLPNADR  
AFIGHFGVLOLPARKPEYKVGVEVIFPGANGKNSAEQHOTELTVKRYSDADV  
QIMYDASWEPREYVHKGLISNATIIYHIDTPAIPGIRIRIRIGHNRKQELLRVAV  
ILAFGISSEFEVYVT"

BASE COUNT 888 a 760 c 714 g 746 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 2049; DB 10; Length 3108;

Mismatches 2049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1..TTCAGTGGCTACTACATTGGCGTTGGAGAGCGGATTCACAGACAGTGTACATATC  
948 TTCAGTGGCTACTACATTGGCGTTGGAGAGCGGATTCACAGACAGTGTACATATC  
61 AATTGATGGGCTATGGCAAAATGGCCAGAAATGCACGGGGTCTCTCACAGGCTGTTG  
1008 AATTGATGGGCTATGGCAAAATGGCCAGAAATGCACGGGGTCTCTCACAGGCTGTTG  
121 ACCCGGCTTTATCTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTGGAGCGG  
1068 ACCCGGCTTTATCTGGCGGATCCAGATGGGTCAAAATGCAATGGCATTTGGAGCGG  
1128 GAATATGATGATATTTCCCAAGCAGTGAAGTGGAGGTCCTGAAGACTAGAGATAA  
241 TATGGCTCTCTGATCGAAGACAAATGTTATCTGAGTGCATTCACACACTCTGGC  
1188 TATGGCTCTCTGATCGAAGACAAATGTTATCTGAGTGCATTCACACACTCTGGC  
301 CCAGAGGGTTTTTCCCAATATACACTGTATATCTGCGCAGGAGGATTCAGCAACGG  
1248 CCAGAGGGTTTTTCCCAATATACACTGTATATCTGCGCAGGAGGATTCAGCAACGG  
361 ACCTTTCACTACATAGTCTCTGGATCATGAAGAGCATTTGATAGCTCACACAATCTT  
1308 ACCTTTCACTACATAGTCTCTGGATCATGAAGAGCATTTGATAGCTCACACAATCTT

Db 1367

421 AAACCAAGCAAAATCTTTTCAACAAGAAATGTTGCTAATGTGCATGCAACGAGAC 480  
1368 AAACCAAGCAAAATCTTTTCAACAAGAAATGTTGCTAATGTGCATGCAACGAGAC 1427  
481 CCCTCCTTCTTACCTTGATCCACAGTGAAGAGAGCAAGGATATCTTCAACACAGAC 540  
1428 CCCTCCTTCTTACCTTGATCCACAGTGAAGAGAGCAAGGATATCTTCAACACAGAC 1487  
541 AAGGAAATGCTGATCTTGAAGTGTGAATTTGAATGAGAAAGACTTGGCTTATACAC 600  
1488 AAGGAAATGCTGATCTTGAAGTGTGAATTTGAATGAGAAAGACTTGGCTTATACAC 1547  
601 TGGTTTGGCTATCCACCCCTGAGCATGAACCATACCTTGTGATAGTACAT 660  
1548 TGGTTTGGCTATCCACCCCTGAGCATGAACCATACCTTGTGATAGTACAT 1607  
661 ATGGGCTATGGGCTTACCTTTTGAAGAAAGAAACAAAGGCTATGCTGGAGAC 720  
1608 ATGGGCTATGGGCTTACCTTTTGAAGAAAGAAACAAAGGCTATGCTGGAGAC 1667  
721 GGACCGTTTGTAGCAGGCTTTGCTTCATCAATATCTCGAGAGCTGCACCAACATCTT 780  
1668 GGACCGTTTGTAGCAGGCTTTGCTTCATCAATATCTCGAGAGCTGCACCAACATCTT 1727  
781 GGCCCGCATTTGTCTACACACAGGAGATTTGTGACACGACAGACACCTGTCCCAAC 840  
1728 GGCCCGCATTTGTCTACACACAGGAGATTTGTGACACGACAGACACCTGTCCCAAC 1787  
841 GGTGGGCGCATGATGCGATGCGACAGCGGACCTGGCAAGCATGTTGAGAGCACACAC 900  
1788 GGTGGGCGCATGATGCGATGCGACAGCGGACCTGGCAAGCATGTTGAGAGCACACAC 1847  
901 AATTATAGCAGGATCATCTATCAGAAAGCAAGAGCTGTATGCTGTCCCTCCAGAG 960  
1848 AATTATAGCAGGATCATCTATCAGAAAGCAAGAGCTGTATGCTGTCCCTCCAGAG 1907  
961 GTGACCGGCGCATGCTTGCACACTCACCGTGGGTAACATACAGATGTGAGCTCCAG 1020  
1908 GTGACCGGCGCATGCTTGCACACTCACCGTGGGTAACATACAGATGTGAGCTCCAG 1967  
1021 CTCATATGCCACACACAGTGAAGAGCTGTAAACCTGGCCGTGCTACAGTTTCCGCA 1080  
1968 CTCATATGCCACACAGTGAAGAGCTGTAAACCTGGCCGTGCTACAGTTTCCGCA 2027  
2028 GCACAAATGATGAGATTTGGGGCTCAATATTTAACAGAGCAATACGGAAGGGATCCA 2087  
1081 GGCACAAATGATGAGATTTGGGGCTCAATATTTAACAGAGCAATACGGAAGGGATCCA 1140  
2087 GGCACAAATGATGAGATTTGGGGCTCAATATTTAACAGAGCAATACGGAAGGGATCCA 2087  
1141 TTCTGGGACACTTTCGGGACCAAGCTCTGGGAAAAACATCTGGAAGAGTTGTAAAGTGT 1200  
2088 TTCTGGGACACTTTCGGGACCAAGCTCTGGGAAAAACATCTGGAAGAGTTGTAAAGTGT 2147  
1201 CAGAAACCAACCAATCTGCTTACAGATGAGAGCTGACATACCAATCTTGGGCA 1260  
2148 CAGAAACCAACCAATCTGCTTACAGATGAGAGCTGACATACCAATCTTGGGCA 2207  
1261 CCAGATATTTGATGATGTTGATGTTTACCGTTGGGTCCTTGGCCATAGCTATCCCT 1320  
2208 CCAGATATTTGATGATGTTGATGTTTACCGTTGGGTCCTTGGCCATAGCTATCCCT 2267  
1321 GGGGAATTAACAACATGTCGGGAGAGAAATTTCTGTAGAGCAATTAAGAAATTTGCA 1380  
2268 GGGGAATTAACAACATGTCGGGAGAGAAATTTCTGTAGAGCAATTAAGAAATTTGCA 2327  
1381 CTTTATGGATGAAGATATGACCTGTTGTTATCGAGAGCTTAAGCAATGTTTATACAT 1440  
2328 CTTTATGGATGAAGATATGACCTGTTGTTATCGAGAGCTTAAGCAATGTTTATACAT 2387  
1441 TACATTACCAATATGAGAAATATACAGGCTACGGGTACGAGGACAGATCTTACATCTAT 1500  
2388 TACATTACCAATATGAGAAATATACAGGCTACGGGTACGAGGACAGATCTTACATCTAT 2447  
1501 GGACCAACACCCCTGTCTGCATACATCCAACTCTTACAGAGACCTTGTAAAGCAATTTGCT 1560

```

Db      2448  GGACACACACACCCTGTCTGACATACATCCAACTCTCAGAGACTTGCTAAGCAATTCCT
Oy      1561  ACGGACACAGTAGGCAACATAGAGCGTGTCCCGGCTCCATCTCTCAAAATCTATA
Db      2508  ACGGACACAGTAGGCAACATAGAGCGTGTCCCGGCTCCATCTCTCAAAATCTATA
Oy      1621  GCTTCACCTTATTCCTAATATTCGGATAGAGCAACCAATTTGGGATGTC
Db      2568  GCTTCACCTTATTCCTAATATTCGGATAGAGCAACCAATTTGGGATGTC
Oy      1681  TTGACAGCCAGAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATATTTGAGGC
Db      2628  TTGACAGCCAGAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATATTTAGGC
Oy      1741  GCTAACCCAAAGAAATTCAGAGAGACAGACCACTCAACCTTCCTCAGTGGAGAAA
Db      2688  GCTAACCCAAAGAAATTCAGAGAGACAGACCACTCAACCTTCCTCAGTGGAGAAA
Oy      1801  TACGAGACTCTGTAGCTGAGTGGAGATTAATGTATACAGATCCCTCGGAGAGAG
Db      2748  TACGAGACTCTGTAGCTGAGTGGAGATTAATGTATACAGATCCCTCGGAGAGAG
Oy      1861  TTTTATTTGGCAAAAGAAATCTGGGTCTGAGCAATGCAACAATATACCTGGCATATTC
Db      2808  TTTTATTTGGCAAAAGAAATCTGGGTCTGAGCAATGCAACAATATATCTGGCATATTC
Oy      1921  GATCTGCTACCTCCGGAATCTAGCAATATTAATTTTGGACAAATCCGAGAGAGAA
Db      2868  GATCTGCTACCTCCGGAATCTAGCAATATTAATTTTGGACAAATCCGAGAGAGAA
Oy      1981  CTCTGAAACCCGCTGTCTATCTAGCATTTGAAGAAATTTCTCTCTTTGAAGTGTG
Db      2928  CTCTGAAACCCGCTGTCTATCTAGCATTTGAAGAAATTTCTCTCTTTGAAGTGTG
Oy      2041  ACTACTAG 2049
Db      2988  ACTACTAG 2996

RESULT 2
AB037181 4835 bp mRNA linear ROD 08-JUN-2001
LOCUS      Mus musculus BCDase mRNA for neutral ceramidase, complete cds.
DEFINITION
AB037181
ACCESSION  AB037181.1 GI:7576310
VERSION     neutral ceramidase.
KEYWORDS    Mus musculus tissue;_1lb:brain cDNA to mRNA.
SOURCE      Mus musculus
ORGANISM    Mus musculus
REFERENCE   1 (sites)
AUTHORS     Tani,M., Okino,N., Mori,K., Tanigawa,T., Izu,H. and Ito,M.
TITLE       Molecular cloning of the full-length cDNA encoding mouse neutral
ceramidase. A novel but highly conserved gene family of
neutral/alkaline ceramidases
JOURNAL     J. Biol. Chem. 275 (15), 11229-11234 (2000)
MEDLINE     20219171
REFERENCE   2 (bases 1 to 4835)
AUTHORS     Okino,N. and Ito,M.
TITLE       Direct Submission
JOURNAL     Submitted (17-JAN-2000) Makoto Ito, Kyushu University, Graduate
School of Bioresource and Bioenvironmental Sciences, Department of
Bioscience and Biotechnology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka
812-8581, Japan (E-mail:makoto@agr.kyushu-u.ac.jp,
Tel:81-92-642-2900, Fax:81-92-642-2900)
FEATURES
source
Location/Qualifiers
1..4835
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_1lb="Brain"
1..4835
/gene="BCDase"

```

```

CDs      441..2711
/gene="BCDase"
/codon_start=1
/product="neutral ceramidase"
/protein_id="BA94546.1"
/db_xref="GI:7576311"
/translation="MAKRTFTEAEFLIFLIVMTVVALITLLEFVSGTIEHKDS
GNHMFSTLSTGTTQPPPTQFPNPSRPNRSGYIGVGRDQGOVSDINMGKN
GONARGLITRLPSAEFLITADDDSNRMAFVEVELCMISORLEKLEKSYLR
KDNVILSLHHSRPFQYTLITLASEGSRNRFQYIVSGIKMSIDATNINLPRK
IFINGNVANOVIKNSPSSYLINPQSERARYSSNDRKEMVLKLDLNGEDGLTSMF
AIHPVSMNNSHNFVNSDMQYAAALFEDEKRGYLPQGGPVPVAGPSSNLDVSNIL
GPHCVNTEGSCDNDKSTCPNGSPSKWASGPGODFESTHIGRIITXOKAELVAS
OEVTGPVLAHOMVMTDVSVOLNHTVCKCPALGSGYFPAAGTIDGSGNITDGT
EGDPWMDITRDLGKPESEIVECKPKPIILHSGELIIPHPMODIYDVOIVYGL
ALAIHSELITMSGRPREAIKKERALKGMDMYVINGLSNVYTHYITTYEVOAOR
YEASTIGETHLSAYIDFERDLAAITDVAANSSGPEPPFNLIASLIPNADR
APIGHFDVLDPAKPEYRVEVEVIVGANPNKSNQHOHFLIVELYKESVDVAD
QIMVNDASMETRFYWHKGIILGISNATYIWHIPDRVAYPQIYRIRYFGRKQELKPAV
ILAEFGISPEEVYT"
BASE COUNT 1460 a 996 c 1048 g 1331 t
ORIGIN
Query Match 99.9%; Score 2047.4; DB:10; Length 4835;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2048; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy      1  TTCAGTGGCTACTACATTTGGCGTGTGGAGAGCGGATTCACAGACGAAGTGCATATTC
Db      663  TTCACTGCTACTACATTTGGCGTGTGGAGAGCGGATTCACAGACGAAGTGCATATTC
Oy      61  AATTGATGGCGTATGAGCAAAATGCGCAGAAATGCGAGGGGTCTCCACAGAGTGTTC
Db      723  AATTGATGGCGTATGAGCAAAATGCGCAGAAATGCGAGGGGTCTCCACAGAGTGTTC
Oy      121  AGCGTGTCTTATCTTGGCGGATCCAGATGGGTCAATCGAATGCAATTTGTGAGCGTG
Db      783  AGCGTGTCTTATCTTGGCGGATCCAGATGGGTCAATCGAATGCAATTTGTGAGCGTG
Oy      181  GAACATATGTATGATTTCCCAACGACTGAGGTGGAGGTCTCGAAGAGCTAGAGTAA
Db      843  GAACATATGTATGATTTCCCAACGACTGAGGTGGAGGTCTCGAAGAGCTAGAGTAA
Oy      241  TATGCTCTCTGTATCGAAGAGCAATGTTATCTGATGCGCATTCACACACACTCTGCG
Db      903  TATGCTCTCTGTATCGAAGAGCAATGTTATCTGATGCGCATTCACACACACTCTGCG
Oy      301  CCAGCAGGGTTTTTCCAAATATACACTATATATCTGCCAGCAGGAGATTCAGCAACGG
Db      963  CCAGCAGGGTTTTTCCAAATATACACTATATATCTGCCAGCAGGAGATTCAGCAACGG
Oy      361  ACCCTTCAGTACATAGTCTCGGAGTCAATGAAAGACATTTGATATGCTCACCAATCTT
Db      1023  ACCCTTCAGTACATAGTCTCGGAGTCAATGAAAGACATTTGATATGCTCACCAATCTT
Oy      421  AAACGAGCAAAATTTTATCAACAAGAAATGTGCTAATGTGCATATCAACGAGC
Db      1083  AAACGAGCAAAATTTTATCAACAAGAAATGTGCTAATGTGCATATCAACGAGC
Oy      481  CCCTCTCTTACCTTCTGAATCCAGAGTGCAGAGAGCAAGATATTTCAACACAGAC
Db      1143  CCCTCTCTTACCTTCTGAATCCAGAGTGCAGAGAGCAAGATATTTCAACACAGAC
Oy      541  AAGGAATGTCGCTTAAACTGTGATTTGAATGAGAAAGATTTGGCTTTTCAGC
Db      1203  AAGGAATGTCGCTTAAACTGTGATTTGAATGAGAAAGATTTGGCTTTTCAGC
Oy      601  TGGTTGGCATTCCACCCGCTGAGCATGACAATATGACCAACTTTGTAATAGTACAT
Db      1263  TGGTTGGCATTCCACCCGCTGAGCATGACAATATGACCAACTTTGTAATAGTACAT
Oy      661  ATGGGCTATGGCGCTTACCTTTTGAAGAGAAAGAAAGAGCTATCTGCTGACAG

```

Db	1333	ATGGGCTATGCGGGCTTACCTTTTGGACAGAAAAAGAACAAAGGCTATCTGCTGGACAG	1382
Oy	721	GGACCGTTGTAGCAGCGCTTGTCTTATCAATTAATCTCGAGACGCTGTACCCAACTTCTT	780
Db	1383	GGACCGTTGTAGCAGCGCTTGTCTTATCAATTAATCTCGAGACGCTGTACCCAACTTCTT	1442
Oy	781	GGCCCGCATTTGTGTCACACAGGGGAGTCTTGTGCAMCGACCAAGAGCACTGTGCCAAC	840
Db	1443	GGCCCGCATTTGTGTCACACAGGGGAGTCTTGTGCAMCGACCAAGAGCACTGTGCCAAC	1502
Oy	841	GGTGGGCGCTAGCATGTGCGATGGCCAGCGGACCTGTGACAAAGCATGTTTGAAGACACAC	900
Db	1503	GGTGGGCGCTAGCATGTGCGATGGCCAGCGGACCTGTGACAAAGCATGTTTGAAGACACAC	1562
Oy	901	ATTATAGACCGATATCTATTCAGAGGGCCAAAGAGCTGTATGCCCTGTCCCGAGAG	960
Db	1563	ATTATAGACCGATATCTATTCAGAGGGCCAAAGAGCTGTATGCCCTGTCCCGAGAG	1622
Oy	961	GTGACCGGCGCAGTGTGCGACGTCACACAGTGGGTGAACATGACAGATGTAGGCTCCAG	1020
Db	1623	GTGACCGGCGCAGTGTGCGACGTCACACAGTGGGTGAACATGACAGATGTAGGCTCCAG	1682
Oy	1021	CTCAATGGCACACACACAGTAAGAAGCTGTAAACCTGGCCCTGAGCTACAGTTTGGCCGA	1080
Db	1683	CTCAATGGCACACACACAGTAAGAAGCTGTAAACCTGGCCCTGAGCTACAGTTTGGCCGA	1742
Oy	1081	GGCAACAATTGTATGGAGTTTGGGGCTCTCAATATTCACACAGGGAACTACGGAAGGGATCCA	1140
Db	1743	GGCAACAATTGTATGGAGTTTGGGGCTCTCAATATTCACACAGGGAACTACGGAAGGGATCCA	1802
Oy	1141	TTCTGGGACACTCTCTTGGGACCACTCTTGGGAAAACCATCTGAAGAGATTGTAGAGT	1200
Db	1803	TTCTGGGACACTCTCTTGGGACCACTCTTGGGAAAACCATCTGAAGAGATTGTAGAGT	1862
Oy	1201	CAGAAACCCAACAACCAATCTGCTTACAGTGGAGAGCTGACGATACCAATCTTTGGCAA	1260
Db	1863	CAGAAACCCAACAACCAATCTGCTTACAGTGGAGAGCTGACGATACCAATCTTTGGCAA	1922
Oy	1261	CCAGATATTTGTGATGTTCAAGATTTGACGTTTGGGTCCTTGGCCATAGCTGTATCCCT	1320
Db	1923	CCAGATATTTGTGATGTTCAAGATTTGACGTTTGGGTCCTTGGCCATAGCTGTATCCCT	1982
Oy	1321	GGGGAATTAACAACCATGTGCGGACGAAAGATTTTGTGAGGCAATTTAAAAAGAATTTGCCA	1380
Db	1983	GGGGAATTAACAACCATGTGCGGACGAAAGATTTTGTGAGGCAATTTAAAAAGAATTTGCCA	2042
Oy	1381	CTTTATGGGATGAAGATATGACCGTGTATTTCAGAGCTTAAGCAATGTTTATACAT	1440
Db	2043	CTTTATGGGATGAAGATATGACCGTGTATTTCAGAGCTTAAGCAATGTTTATACAT	2102
Oy	1441	TACATTAACCATATATGAAGAATTACCAAGGCTCAGCGGTAAGAGGAGCATCTACATCTAT	1500
Db	2103	TACATTAACCATATATGAAGAATTACCAAGGCTCAGCGGTAAGAGGAGCATCTACATCTAT	2162
Oy	1501	GGACACACACACCTGTCTGCATTAATCTCAACTCTTCAGAGACCTTGTCTAAGGCAATTGCT	1560
Db	2163	GGACACACACACCTGTCTGCATTAATCTCAACTCTTCAGAGACCTTGTCTAAGGCAATTGCT	2222
Oy	1561	ACGGACACAGTAGGCAACATGAGAGAGTGGTCCGAGGCTCCATTCTTCAAAATCTATA	1620
Db	2223	ACGGACACAGTAGGCAACATGAGAGAGTGGTCCGAGGCTCCATTCTTCAAAATCTATA	2282
Oy	1621	GCTTCACATTATCTTAATATTTGCGGATATGAGACACCAATTTGGCAAAATTTTGGGGATGTC	1680
Db	2283	GCTTCACATTATCTTAATATTTGCGGATATGAGACACCAATTTTGGGGATGTC	2342
Oy	1681	TTGCAAGCCAGCAAAACCTGAATACAGAGTGGGAGAGTGTGGAAGTTATATTTGTAGGC	1740
Db	2343	TTGCAAGCCAGCAAAACCTGAATACAGAGTGGGAGAGTGTGGAAGTTATATTTGTAGGC	2402
Oy	1741	GCTAAACCAAAAGATTTGACGAGAAACCAAGACCATCAAAACCTTCTCAGCTGTGGAGAA	1800
Db	2403	GCTAAACCAAAAGATTTGACGAGAAACCAAGACCATCAAAACCTTCTCAGCTGTGGAGAA	2462

REFERENCE	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE	FEATURES
1801	TACAGAGACCTGTAGCTGACATGGCAGATATGTATTAACGATTCGCTCTGGAGACGAG	2456 bp	AB057433	GI:14701597	Rattus norvegicus kidney cDNA to mRNA.	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2463	TACGAGACCTGTAGCTGACATGGCAGATATGTATTAACGATTCGCTCTGGAGACGAG	2522	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1861	TTTATTGGCAACGAGTACTGGTCTGAGCAATGCACAACTATTAAGTATTC	1920	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1921	GATACCTGCTTACCTGGATATACAGATTAAGATATTTTGGACCAATCGAGACAGAA	1960	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2583	GATACCTGCTTACCTGGATATACAGATTAAGATATTTTGGACCAATCGAGACAGAA	2642	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1981	CTTCTGAAACCCGCTGATACATGATTAAGAAATTTCTCTTTTGAAGTTGTC	2040	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2643	CTTCTGAAACCCGCTGATACATGATTAAGAAATTTCTCTTTTGAAGTTGTC	2702	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2041	ACTACTTAG 2049		AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2703	ACTACTTAG 2711		AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1801	TACAGAGACCTGTAGCTGACATGGCAGATATGTATTAACGATTCGCTCTGGAGACGAG	1860	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2463	TACGAGACCTGTAGCTGACATGGCAGATATGTATTAACGATTCGCTCTGGAGACGAG	2522	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1861	TTTATTGGCAACGAGTACTGGTCTGAGCAATGCACAACTATTAAGTATTC	1920	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1921	GATACCTGCTTACCTGGATATACAGATTAAGATATTTTGGACCAATCGAGACAGAA	1960	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2583	GATACCTGCTTACCTGGATATACAGATTAAGATATTTTGGACCAATCGAGACAGAA	2642	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1981	CTTCTGAAACCCGCTGATACATGATTAAGAAATTTCTCTTTTGAAGTTGTC	2040	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2643	CTTCTGAAACCCGCTGATACATGATTAAGAAATTTCTCTTTTGAAGTTGTC	2702	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2041	ACTACTTAG 2049		AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2703	ACTACTTAG 2711		AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1801	TACAGAGACCTGTAGCTGACATGGCAGATATGTATTAACGATTCGCTCTGGAGACGAG	1860	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2463	TACGAGACCTGTAGCTGACATGGCAGATATGTATTAACGATTCGCTCTGGAGACGAG	2522	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1861	TTTATTGGCAACGAGTACTGGTCTGAGCAATGCACAACTATTAAGTATTC	1920	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1921	GATACCTGCTTACCTGGATATACAGATTAAGATATTTTGGACCAATCGAGACAGAA	1960	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2583	GATACCTGCTTACCTGGATATACAGATTAAGATATTTTGGACCAATCGAGACAGAA	2642	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
1981	CTTCTGAAACCCGCTGATACATGATTAAGAAATTTCTCTTTTGAAGTTGTC	2040	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2643	CTTCTGAAACCCGCTGATACATGATTAAGAAATTTCTCTTTTGAAGTTGTC	2702	AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2041	ACTACTTAG 2049		AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456
2703	ACTACTTAG 2711		AB057433	GI:14701597	Rattus norvegicus	Rattus norvegicus	Submitted (14-MAR-2001)	1. .2456

SVANQJMHNDASHMETRPFYHKGVLGSLNATIHWHIPDLPVRYIRYFGRKROEL  
LKPXVILAFEBISSPFEIVTT"  
BASE COUNT 704 a 612 c 562 g 578 t  
ORIGIN

Query Match 89.5%; Score 1833; DB 10; Length 2456;  
Best Local Similarity 93.4%; Pred. No. 0;  
Matches 1914; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

1 TTCAGTGGCTACATTTGGCGTTGGGAGCGGATGGACAGACAAATGTCAGATATC 60  
|||||  
307 TTCATGGCTACTACATTTGGTTGGGAGCTGACCTGACCGAACAATGTCAGATATC 366  
|||||  
61 AATTGATGGCTATGGCAAAAATGGCCAGATGACGAGGCTCTCTCCGCCAGCTGTC 120  
|||||  
367 AATTGATGGCTATGGCAAAAATGGCCAGACAGAGGCTCTCTCCGCCAGCTGTC 426  
|||||  
121 AGCGGTCTTTTATCTTGGCGGATCCAGATGGTCAATGCAATGGCATTTTGTAGCGTG 180  
|||||  
427 AGCGGTCTTTTATCTTGGCGGATCCAGATGGTCAATGCAATGGCATTTTGTAGCGTG 486  
|||||  
181 GAATATGATGATTTCCCAAGCAGTGGTTGGAGGCTCGAAGAGCTGAGAGATAA 240  
|||||  
487 GAATATGATGATTTCCCAAGCAGTGGTTGGAGGCTCGAAGAGCTGAGAGATAA 546  
|||||  
241 TATGGCTCTCTGATTCGAGAGACAAATGTTATCCGTGAGTTCATTCACACACTGCGC 300  
|||||  
547 TATGGCTCTCTGATTCGAGAGACAAATGTTATCCGTGAGTTCATTCACACACTGCGC 606  
|||||  
301 CCACAGAGGTTTTCATATACACTCTATATCTCCGACGAGAGGATTCAGCAACCGG 360  
|||||  
607 CCACAGAGGTTTTCATATACACTCTATATCTCCGACGAGAGGATTCAGCAACCGG 666  
|||||  
361 ACCTTATGATGATGCTGCGGATCATGAAGAGCATGATATAGTTCACACAAATCTT 420  
|||||  
667 ACCTTATGATGATGCTGCGGATCATGAAGAGCATGATATAGTTCACACAAATCTT 726  
|||||  
421 AAACAGGCAAAAATCTTTATCAACAAGAAATGTTGCTATGTCAGATCAACGGAAT 480  
|||||  
727 AAACAGGCAAAAATCTTTATCAACAAGAAATGTTGCTATGTCAGATCAACGGAAT 786  
|||||  
481 CCTCTCTCTTACCTTGAATCCAGAGTCCAGAGAGAGAGATATCTTCAACACAGAC 540  
|||||  
787 CCTCTCTCTTACCTTGAATCCAGAGTCCAGAGAGAGAGATATCTTCAACACAGAC 846  
|||||  
541 AAGGAATGCGGCTGTAACCTGTAATGTAATGTAAGAGAGAGTGGCTTATACGC 600  
|||||  
847 AAGGAATGCGGCTGTAACCTGTAATGTAATGTAAGAGAGAGTGGCTTATACGC 906  
|||||  
601 TGGTTGGCATCCACCCGTCGAGCATGAACATAGCAACACTTGTGATATGACAAAT 660  
|||||  
907 TGGTTGGCATCCACCCGTCGAGCATGAACATAGCAACACTTGTGATATGACAAAT 966  
|||||  
661 ATGGGCTATGCGGCTTACCTTTTGAACAAGAAAACAAGGCTATCTGCGTGGACAG 720  
|||||  
967 ATGGGCTATGCGGCTTACCTTTTGAACAAGAAAACAAGGCTATCTGCGTGGACAG 1026  
|||||  
721 GGACCGTTTGTAGAGGCTTGTCTTCAATCTCGAGACGTCGTCACCAATCTT 780  
|||||  
1027 GGACCGTTTGTAGAGGCTTGTCTTCAATCTCGAGACGTCGTCACCAATCTT 1086  
|||||  
781 GGCCCGCATGTCACACAGGAGGAGTCTTGTGACAAAGCAAGAGACCTGTGCCAAC 840  
|||||  
1087 GGCCCGCATGTCACACAGGAGGAGTCTTGTGACAAAGCAAGAGACCTGTGCCAAC 1146  
|||||  
841 GGTGGGCTTACATGTCATGCGCAGCGAGCTGGACAGACAAATGTTGAGAGACACAC 900  
|||||  
1147 GGTGGGCTTACATGTCATGCGCAGCGAGCTGGACAGACAAATGTTGAGAGACACAC 1206  
|||||  
901 AATTAGAGGATCATCTATCAAGAGCAAGAGAGCTGATGCTCTGCTCCAGAGAG 960  
|||||  
1207 AATTAGAGGAGGCTCATCTATCAAGAGCAAGAGAGCTGATGCTCTGCTCCAGAGAA 1266  
|||||

961 GTGACCGGCCACAGTCTTGCACATCCACAGTGGGTGAACATAGAGATGTGAGCGTCAG 1020  
|||||  
1267 GTGACCGGCCACAGTCTTGCACATCCACAGTGGGTGAACATAGAGATGTGAGCGTCAG 1326  
|||||  
1021 CTCATGTCACACACAGTGAAGAGCTGTAAACCTGCTCGGCTACAGTTTGGCCGA 1080  
|||||  
1327 CTCATGTCACACACAGTGAAGAGCTGTAAAGCGCCGCTACAGTTTGGCCGA 1386  
|||||  
1081 GGCACAAATGATGAGTGGGCTGCAATTTACAGAGGAGACTACGGAAGGGATGCCA 1140  
|||||  
1387 GGCACAAATGATGAGTGGGCTGCAATTTACAGAGGAGAGCAAGAGGAGATCTA 1446  
|||||  
1141 TTCTGGGACACTCTTGGGACCAAGCTTGGGAAACCATCTGAAGATTTGAGAGTGT 1200  
|||||  
1447 TTCTGGGACACTCTTGGGACCAAGCTTGGGAAACCATCTGAAGATTTGAGAGTGT 1506  
|||||  
1201 CAGAAACCCAAACCAATCTCTCTTACAGTGGAGAGCTGACGATACCATCTTGGCAA 1260  
|||||  
1507 CAGAAACCCAAACCAATCTCTCTTACAGTGGAGAGCTGACGAAACCTCATCTTGGCAA 1566  
|||||  
1261 CCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
|||||  
1567 CCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1626  
|||||  
1321 GGGGAAATTAACACCATGTCGGGAGAGATTTGCGTGAAGCAATTAAGAAATTTGCA 1380  
|||||  
1627 GGGGAAATTAACACCATGTCGGGAGAGATTTGCGTGAAGCAATTAAGAAATTTGCA 1686  
|||||  
1381 CTTTATGGGATGAAGATATGACCGTGTGTATGCGAGGTCTTAACGATGTTTATACAT 1440  
|||||  
1687 CTTTATGGGATGAAGATATGACCGTGTGTATGCGAGGTCTTAACGATGTTTATACAT 1746  
|||||  
1441 TACATTTACCATATGAAGATATGACCGTGTGTATGCGAGGTCTTAACGATGTTTAT 1500  
|||||  
1747 TACATTTACCATATGAAGATATGACCGTGTGTATGCGAGGTCTTAACGATGTTTAT 1806  
|||||  
1501 GGACACACACACCGTCTGCTGATACATCCACTTTCAGAGACTTGTGTAAGCAATTTGCT 1560  
|||||  
1807 GGACACACACACCGTCTGCTGATACATCCACTTTCAGAGACTTGTGTAAGCAATTTGCT 1866  
|||||  
1561 ACGGACACATGACCAACATGACAGTGGTCCGAGCTCCATTTCTTCAAAAATCTAATA 1620  
|||||  
1867 ACGGACACATGACCAACATGACAGTGGTCCGAGCTCCATTTCTTCAAAAATCTAATA 1926  
|||||  
1621 GCTTACTTATTTCTTAATTTGCGGATGAGACCAATTTGCAAAATTTTGGGAGTGC 1680  
|||||  
1927 GCTTACTTATTTCTTAATTTGCGGATGAGACCAATTTGCAAAATTTTGGGAGTGC 1986  
|||||  
1681 TTGACGCGCAGCAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATTTGTAGGC 1740  
|||||  
1987 TTGACGCGCAGCAAAACCTGAATACAGAGTGGAGAAAGTGTGAAGTTATTTGTAGGC 2046  
|||||  
1741 GCTAACCCAAAGAAATTCAGCAGAGAACCAACCATCAACCTTCTCATCTGTGGAGAA 1800  
|||||  
2047 GCTAACCCAAAGAAATTCAGCAGAGAACCAACCATCAACCTTCTCATCTGTGGAGAA 2106  
|||||  
1801 TACGAGACCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1860  
|||||  
2107 TACGAGACCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2166  
|||||  
1861 TTATTATTGGCAAAAGAAATTTGCTGCTGAGCAATGCAACAAATTTGCTGAGATTTCCA 1920  
|||||  
2167 TTATTATTGGCAAAAGAAATTTGCTGCTGAGCAATGCAACAAATTTGCTGAGATTTCCA 2226  
|||||  
1921 GATACGCTTACCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1980  
|||||  
2227 GATACGCTTACCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2286  
|||||  
1981 CTTCTGAACCCGCTGTCATCTGATGCAATTTGAAGAAATTTCTCTTTTGAAGTGTG 2040  
|||||  
2287 CTTCTGAACCCGCTGTCATCTGATGCAATTTGAAGAAATTTCTCTTTTGAAGTGTG 2346  
|||||  
2041 ACTACTTAG 2049



```

Db      2347 ACTACTAG 2355

RESULT 4
E50433
LOCUS   E50433          2283 bp      DNA          linear      PAT 31-JAN-2002
DEFINITION Novel protein.
ACCESSION E50433
VERSION   E50433.1 GI:18633518
KEYWORDS  JP 2001057890-A/1.
SOURCE    Rattus norvegicus.
ORGANISM  Rattus norvegicus.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE 1 (bases 1 to 2283)
AUTHORS  Omori, A. and Ito, M.
TITLE     Novel protein.
JOURNAL   Patent: JP 2001057890-A 1 06-MAR-2001;
           MITSUBISHI CHEM CORP
COMMENT   OS Rattus norvegicus (rat)
           PN JP 2001057890-A/1
           PD 06-MAR-2001
           PF 23-AUG-1999 JP 1999235218
           PR
           PI AKIRA OMORI, MAKOTO ITO
           PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/80, PC
           CC C12N15/00,C12N5/00
           FT
           FH Key Location/Qualifiers
           CDS (1)..(2283).
           Location/Qualifiers
           1..2283
           /organism="Rattus norvegicus"
           /db_xref="taxon:10116"
BASE COUNT 641 a 584 c 524 g 534 t
ORIGIN
Query Match      89.3%; Score 1830; DB 6; Length 2283;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1911; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```

```

Db      658 AAACGGGCAAGTCCTTATCAACAAAGGAATGTTGCTAATGTCAGATCAACCGCACT 717
QY      481 CCCCTCTTACCTTCTGAATCCACAGTCAGAGAGAGCAAGGTATTTCTTCAACACAGAC 540
Db      718 CCCCTCTTACCTTCTGAATCCACAGTCAGAGAGAGCAAGGTATTTCTTCAACACAGAC 777
QY      541 AAGGAATGCTGCTCTTGAAGTGTGATTTGAATGGAAGACTTGGGTCTATTCAC 600
Db      778 AAGGAATGCTGCTCTTGAAGTGTGATTTGAATGGAAGACTTGGGTCTATTCAC 837
QY      601 TGGTTGGCATCCACCCCGTGAGCAATGAAACATACCACTTGTGATGTGACAT 660
Db      838 TGGTTGGCATCCACCCCGTGAGCAATGAAACATACCACTTGTGATGTGACAT 897
QY      661 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAAGCAAGCAATGCTGCTGAGAC 720
Db      898 ATGGGCTATGCGGCTTACCTTTTGAAGCAAGAAAGCAAGCAATGCTGCTGAGAC 957
QY      721 GGACCGTTGTAGCAGGCTTGTCTTATCAATATCTGAGAGAGCTGACCCAACTTCT 780
Db      958 GGACCGTTGTAGCAGGCTTGTCTTATCAATATCTGAGAGAGCTGACCCAACTTCT 1017
QY      781 GGCCCGCATTTGTCAACAGAGGGAGTGTGTGAACAAGAGAGCACTGTCCCAAC 840
Db      1018 GGCCCGCATTTGTCAACAGAGGGAGTGTGTGAACAAGAGAGCACTGTCCCAAC 1077
QY      841 GGTGGGCTTATGATGTCATGCGCAGCGGACCTTGCAACAGATGTTGAGACACAC 900
Db      1078 GGTGGGCTTATGATGTCATGCGCAGCGGACCTTGCAACAGATGTTGAGACACAC 1137
QY      901 ATTATAGGAGCGGATATCTATCAGAAAGGCAAGAGCTGTATGCTTGTCTCCAGAG 960
Db      1138 ATTATAGGAGCGGATATCTATCAGAAAGGCAAGAGCTGTATGCTTGTCTCCAGAG 1197
QY      961 GTGACCGGCGGCGGCTTGTCTTCACTACAGAGTGGTGAACATGACAGATGAGCGCTCAG 1020
Db      1198 GTGACCGGCGGCGGCTTGTCTTCACTACAGAGTGGTGAACATGACAGATGAGCGCTCAG 1257
QY      1021 CTCATGCGCACACACAGTGAAGAGTGTAAACCTGCGCTGAGTATGTTGCGCGCA 1080
Db      1258 CTCATGCGCACACACAGTGAAGAGTGTAAACCTGCGCTGAGTATGTTGCGCGCA 1317
QY      1081 GGCACATTTGATGAGTTCGCGGCTTCAATATTACAGAGGAACATACGGAAGGGATCCA 1140
Db      1318 GGCACATTTGATGAGTTCGCGGCTTCAATATTACAGAGGAACATACGGAAGGGATCCA 1377
QY      1141 TTCTGGACACTCTTGGGACAGCTCTTGGGAAACATCTGGAAGAGATTGTAAGTGT 1200
Db      1378 TTCTGGACACTCTTGGGACAGCTCTTGGGAAACATCTGGAAGAGATTGTAAGTGT 1437
QY      1201 CAGAAACCCAAACCAATCTGCTTCAACAGTGAAGAGCTGAGATACCAATCTTGGCAA 1260
Db      1438 CAGAAACCCAAACCAATCTTCAACAGTGAAGAGCTGAGATACCAATCTTGGCAA 1497
QY      1261 CCAGATATTGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 1320
Db      1498 CCAGATATTGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 1557
QY      1321 GGGGAATTAACCAACATGTCGGGAGCAAGATTTGCTGAGGCAATTAACCAATTTGCA 1380
Db      1558 GGGGAATTAACCAACATGTCGGGAGCAAGATTTGCTGAGGCAATTAACCAATTTGCA 1617
QY      1381 CTTTATGGATGAAGATATGACGCTTGTATGCAAGCTGTAAAGCAATTTATACAT 1440
Db      1618 CTTTATGGATGAAGATATGACGCTTGTATGCAAGCTGTAAAGCAATTTATACAT 1677
QY      1441 TACATTACCAATATGAAGAAATATACAGAGTCCAGGCTAGAGGCAAGCAATCTAT 1500
Db      1678 TACATTACCAATATGAAGAAATATACAGAGTCCAGGCTAGAGGCAAGCAATCTAT 1737
QY      1501 GGACCAACACCCCTGCTGATACATCAATCAACTCTTCAAGAGACTTGTAAAGCAATTTGT 1560
Db      1738 GGACCAACACCCCTGCTGATACATCAATCAACTCTTCAAGAGACTTGTAAAGCAATTTGT 1797

```





Db 1176 ATTATAGAGCGCCATGTATACAGAGCAAAAGAACTCTATGCTTGCCTCCAGAG 1235  
Qy 961 GTGACCGGCGCCAGTGTTCAGAGCTCACCAGTGGGTGAACATGACAGATGTGAGCTCCAG 1020  
Db 1236 GTAAAGAGGACACAGCTTTCAGACAGAGTGGGTGATGATGACAGATGTGACCTCTCG 1295  
Qy 1021 CTGATGCGACACACAGAGTGAAGAGCTGTAACTGCCCTGGGTACAGATTTTCCCGCA 1080  
Db 1296 CTCAATTCACACATGATCAAAAACATGTAACACAGATGTGGGTACAGATTTTCCGCT 1355  
Qy 1081 GGCAATTTGATGAGTTCGGGCTCAATATTACAGAGGGAACATGAGGAAGGGATGCA 1140  
Db 1356 GGCATATTGATGAGTGTGGAGGCTCAATTTTACACAGGGGAAAAACAGAGGGATGCA 1415  
Qy 1141 TTCTGGAGACCTCTTGGGACAGCTCTTGGGAAAAACCATCTGAAGAGATTGTAGAGT 1200  
Db 1416 TTTTGGGACCATCTGGGACAGATCTGGGAAAGCCATCTGAAGAAATTAAAGAAATG 1475  
Qy 1201 CAGAAACCCAAACCAATCTGCTTCACAGTGAAGAGCTGACAGATACACATCTTGGCAA 1260  
Db 1476 CATAAACCAAGCCCATCTTCTTCACACCGGAGAACTATCAAAACCTCACCCCTGGCAT 1535  
Qy 1261 CCAGATATTGATGATTCAGATTTGTAACGTTGGTCTTGGCCATAGCTGATCCCT 1320  
Db 1536 CCAGCATTTGATGATTCAGATTTATACCTTGGGTCTTGGCCATAGCTGATCCCT 1595  
Qy 1321 GGGGATTTAGAACCATCTCGGAGCAAGATTTGCTGAGCAATTAAAAAGAAATTTGCA 1380  
Db 1596 GGGGATTTAGAACCATCTCGGAGCAAGATTTGCTGAGCAATTAAAAAGAAATTTGCA 1655  
Qy 1381 CTTATGAGTGAAGATATACCGTCTTATCGAGCTCTAAGCAATTTATACAT 1440  
Db 1656 TCTCTGAGTGAAGATATACCGTCTTATCGAGCTCTAAGCAATTTATACAT 1715  
Qy 1441 TACATTTACATATGAAGATATACAGGCTCAGCGTACGAGAGCATCTACATCTAT 1500  
Db 1716 TACATTTACATATGAAGATATACAGGCTCAGCGTACGAGAGCATCTACATCTAT 1775  
Qy 1501 GGACACACACCCCTGCTGATACATTCACACTCTTCAAGAGACCTTCTAAGCAATTTGCT 1560  
Db 1776 GGACCGGACAGATTTATCTGCTTACATTCAGCTCTTCAAGAACTTCTAAGCAATTTGCT 1835  
Qy 1561 ACAGGACAGTACGCAATGAGAGTGTCCCGAGCTCCATCTCTCAAAAATCTAATA 1620  
Db 1836 ACAGGACAGTACGCAATGAGAGTGTCCCGAGCTCCATCTCTCAAAAATCTAATA 1895  
Qy 1621 GCTTCACTTATTTCTAATATTGCGGATAGAGACACCAATTTCTCAATTTGGGATGTC 1680  
Db 1896 GCTTCACTTATTTCTAATATTGCGGATAGAGACACCAATTTCTCAATTTGGGATGTC 1955  
Qy 1681 TTGCAGCCAGCAAAACCTGAATACAGAGTGGGAGAGTGGTTGAAGTTATATTGTAGGC 1740  
Db 1956 CTGAGCCAGCAAAACCTGAATACAGAGTGGGAGAGTGGTTGAAGTTATATTGTAGGT 2015  
Qy 1741 GCTAACCCAAAGAAATTCAGAGAGAGAACCAACCTCAACCTTCTCACTGTGAGAAA 1800  
Db 2016 GCTAACCCGAAAGAAATTCAGAGAGAGAACCAACCTCAACCTTCTCACTGTGAGAAA 2075  
Qy 1801 TACAGAGACTGTGATGAGTACGAGATTAATGTATTAAGATGCTCCTCGGAGAGAGAG 1860  
Db 2076 TATAGAGCTACTTCAACATCGTGGAGATAGTGTATTAATGATGCTCCTCGGAGAGAG 2135  
Qy 1861 TTTTATTTGGCAAAAGAAATCTGAGTGTGAGCAATGACAAATATTACTGGCATTTTCCA 1920  
Db 2136 TTTTATTTGGCAAAAGAGGCTCCTGGTCTGAGTATTAAGCAAGTGAATTTGCAATTTCCA 2195  
Qy 1921 GATAGTGGTACCTGGAATCTAGAGATTAATTTTGGACACAAATGGAAGAGCA 1980  
Db 2196 GACAGTGGCCAGGCTGGAATCTAGAGATTAATTTTGGACACAAATGGAAGAGCA 2255  
Qy 1981 CTTTGAAGCCGCTGATATAGATTTTGAAGAAATTTCTCTCTTTGAAGTTGTC 2040  
Db 2256 ATTCTGAAGCCTGCTGATATCTTGAAGGCACTTCCCGGCTTTTGAAGTTGTA 2315

Qy 2041 ACTACTAG 2049  
Db 2316 ACTATTAG 2324

RESULT 6  
AF250847  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL

FEATURES  
source

CDS

1..2289  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..2286  
/function="cleaves lipid ceramide into free fatty acid and sphingosine"  
/note="active in neutral to alkaline range"  
/codon\_start=1  
/product="mitochondrial ceramidase"  
/protein\_id="AF250847.1"  
/translation="MSATVVALSLFTTSGTITENHKDLGHFFSTGSPATGSGTA  
AQRSTATSHSTATQSPVPLTPESEPLFQNSGTHVGVGADCTGVADINLM  
GYKSGONAGILTRLYRAFLMAEPDSNRTVEYSIDIGVSOPLRLVLRNLSKY  
GLVLRDNLVILSGTHSGPAGYFQYTFEVLASGEFSNOTOHVYGLISDIADTN  
MKPKITFNKNGVQVQINRSYSYSLQNPQSEBRARYPSNTKEMLVILKMDLNDLC  
LISPAIHVPSKNNNSNHLVNSDNGVYASYLLEQERNKGYLPGQGFVAAPSSNLGDV  
SPNLIGPRICINTGESCDNANSTCPVIGPSMKIAGPGQDMFDSQIIGRAMYORAKEL  
YASAEVETGPLASHOWNDMDVYVNLNTHASKTCRPALGYSFAACTIDGVGLNF  
TQKSGEPFMDTIDQILGKPSSEIKCHRPKPLILHGTGLSKHPHNPPIIDVQIL  
TLGSLAIFAIIGEFTTMSGRRLREVAQEPVASHGQNNNTVVISGLCYTHYITTEE  
YQAKREYASITGPHASAYIQLFRLNAKALADDTVNLISRGPRPEFKQIYDPLP  
SLVDRAPEKRTFDVLPAPKEIRYGEVAEYFVGANKNSYONOTHOTFLVEKTEA  
TSTSWQIVCNDASWETREYVWNGHGLGLISLNATVEMHLPDAPGILYRIRYFHNKRODI  
LKPAVILFEETSPAFEVYTI"

BASE COUNT  
664 a 553 c 515 g 557 t

Query Match 73.6%; Score 1508.2; DB 9; Length 2289;  
Best Local Similarity 83.5%; Pred. No. 0;  
Matches 1711; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

Qy 1 TTCAGTGGCTACTACTATGCGGTGGGAGAGCGGATTCACAGCAAGTGCACATATTC 60  
Db 238 TTCAGTGGCTACTACTATGCGGTGGGAGAGCGGATTCACAGCAAGTGCACATATTC 297

Qy 61 AATTGATGGGCTATGCAAAATGCGAGATGCGAGGAGTCTCTCAGCAGGCTGTTC 120  
Db 298 AATTGATGGGCTATGCGCAAAATGCGAGATGCGAGGAGTCTCTCAGCAGGCTATAC 357

OY	121	AGCGTCGTTTATCTTGGGGGATTCGAGTGGGCAATGGAATGGCATTTGTGAGCGT	180
Db	358	AGTGTGGCTTCAATCATTGGCAGAACTGTAATGGGTCCAAATGAGTGTGTGAGATC	417
OY	181	GAACATATGATATTTCCCAACGACATGAGTTGGAGTCTCGAAGAACATGAGAGTAAA	240
Db	418	GACATATGAGCATGGTATACAAAGGCTAGGCTGGAGCTCTGAAACAGATCGCAGATAAA	477
OY	241	TATGGCTCTGTATCGAAGACAAATGTTATCTGATGGCCATTGCACATACACTTGGC	300
Db	478	TATGGCTCCCTGTACAGAAAGATAATGTACTCTGATGGCACTCAATTCATTCAAGT	537
OY	301	CCAGCAGGGTTTTCCATATACACTATATCTGCCACGAGGATTCAGCACCGG	360
Db	538	CTTCAGAGATATTCACGATATACGGTGTGTGTAATTCACAGTGAAGATTTGAAATCAA	597
OY	361	ACCTTTAGATCATAGTCTCGGGATCATGAAAGCATGTAATATAGTCAACACAAATCTT	420
Db	598	ACCTTTAGCAGCATGTGCTACTGGTATCTTGAAGAGCATTGACATATGACACACAAATATG	657
OY	421	AAACAGGCAAAATCTTATCAACAAAGAAATTTCTATGTGCAATCAACCGAAGC	480
Db	658	AAACAGGCAAAATCTTATCAATAAAGAAATTTGATGGTGGTGCAGATCAACAGAGT	717
OY	481	CCCTCCTTATACCTTCTGATCCACAGTCAGAGAGCAAGATTTCTTCAACACAGAC	540
Db	718	CCGTAATTCCTACCTTCAAAATCCGACATGACAGACCAAGTATCCTTCAAAATACGAC	777
OY	541	AAGAAATATGCTGGCTTGAACCTGTGATTTGAATGAGAAAGCATTTGGGCTATACG	600
Db	778	AAGAAATATGATAGTTTGGAAAATGTGATTTGATATGAGATGACCTTGGGCTTATACG	837
OY	601	TGTTTTCATCCACCCGCTGAGCATGAACATATGCAACATTTGTAATAGTACAAAT	660
Db	838	TGTTTTCATCCACCCGCTGAGCATGAACATATGCAACATTTGTAATAGTACAAAT	897
OY	661	ATGGGCTATGCGGCTTACCTTTTGGACAGAAAGAAACAAAGGCTATGTGCTGACAG	720
Db	898	GTGGGCTATGCTATCTTACCTGCTGACAGAGAAAGAAAGGATATCTACCTGGACAG	957
OY	721	GGACCGTTTGTAGCAGGCTTGTCTTATCCAAATCTGGAGACGCTACCCCAATATCTT	780
Db	958	GGACCGTTTGTAGCAGGCTTGTCTTATCCAAATCTAGAGAGTATGTCTCCCAATATCTT	1017
OY	781	GGCCCGATTTGTCAACAGGAGGATCTTGTGACACGACCAAGACACCTGCTCCCAAC	840
Db	1018	GGACCGATTTGTCAACAGGAGAGTCTCTGTGATATACCCCAATATGCTTGTCCATT	1077
OY	841	GGTGGGCTTACATGTGCATGGCCAGCGACCTGGACAAAGCATGTTTATAGACACAC	900
Db	1078	GGTGGGCTTACATGTGCATGTGCTTAAAGGACCTGACAGGATATGTTTATACAGACAA	1137
OY	901	ATTATAGAGGATCATCTATCTACGAAGGCCAAGGAGTGTATGGCTCTGCTCCAGAG	960
Db	1138	ATTATAGAGGAGGCGCATATGTATCAGAAGCAAGGAATCTATATGCTCTGCTCCAGAG	1197
OY	961	GTGACCGGCGCATGTGTGACATCCACAGTGGGTGAACATGACAGATATGTAGGCTCAG	1020
Db	1198	GTGACCGGCGCATGTGTGACATCCACAGTGGGTGAATGACAGATATGTAGTGTCTGG	1257
OY	1021	CTCAATGCGACACACAGTGAAGAGCTGTAAACCTGCTCCCTGGCTACAGTTTTGCCGA	1080
Db	1258	CTCAATGCGACACATGTGATCAAAAACATGTATTAACACCATTTGGGCTACAGTTTTGCCGT	1317
OY	1081	GGCAGATATGATGAGTGTGGGCTCAATATTTACAGAGGAATCTACGGAAGGAGATCCA	1140
Db	1318	GGCAGATATGATGAGTGTGGGCTCAATATTTTACAGAGGGAAGAAAGAGGAGGATCCA	1377
OY	1141	TTCTTGGGACATCTCTGGGACACACTCTTGGGAAACCATCTGAAGATATGTGAGTGT	1200
Db	1378	TTTCTTGGGACATCTCTGGGACACAGATCTCTGGGAAAGCATCTTGAAGAAATTTAAAGATGT	1437
OY	1201	CAGAAACCAACCAATCTGCTTCAAGTGGAGAGCTGACGATACCAATCCTTTGGCAA	1260

Db	1438	CATAAACCAAGGCCCATCTTCTTCACACGGAGACTATCAAAACCTCACCCCTGGCAT	1497
Oy	1261	CCAGATATATGTTGATGTTTCAGATTTGTTACGGTTGGGTCTTGGCCATAGCGTGCATCCT	1320
Db	1498	CCACACATATGTTGATGTTTCAGATTTATTAACCTTGGGTCTTGGCCATTAAGTGCATCC	1557
Oy	1321	GGGGAATTTAACMACCATGTGGGAGCAAGATTTTCGTGAGCGCAATTTAAAAAATTTGCA	1380
Db	1558	GGGAGATTTCAGACCATGTCTGACGAAGACTTGGAGGCGAGTTCGAAGCAATTTTGA	1617
Oy	1381	CTTATATGGGATGAAGGATATGACCGTGTGTTATGCGAGTCTAAGCATATGTTTATACAT	1440
Db	1618	TCTATATGGGATGCAACATGACGTGTTGTTATTTTCAGGTCTATGCAACGCTATACCAT	1677
Oy	1441	TACATTTCCATATATATGAAGAAATACCAGAGTCAGCGGTACGAGGAGCATCATCAATCAT	1500
Db	1678	TACATTTCCATATATGAAGAAATACCAGAGTCAGCGGTACGAGATATGAGCAGCATATGACATTTAT	1737
Oy	1501	GGACCAACACACCCCTGTCTGCATACATCAACTCTTCAGAGACCTTGGCTAAGGCAATTTGCT	1560
Db	1738	GGACCGGACGCAATATGCTTACATTCAGCTCTTCAGAAACCTTGGCTAAGGCAATTTGCT	1797
Oy	1561	ACGACACACATGACCAACATATGAGAGTGTCCGAGCCTGCATTTCTTCAAAATCTATATA	1620
Db	1798	ACGACACACGATGAGCAACCTGAGAGAGGTCCAAACCTCCCTTTTCAAACAATATATATA	1857
Oy	1621	GCTTCACATTTATTTCTTAATATTTGCGATAGAGACACCAATTTGGCAAAATTTTGGGATGTC	1680
Db	1858	GTTCCATTTAATTTCTAGTATTTGTGATATGAGACCCAAAGGACAGAACTTTCCGGGATGTC	1917
Oy	1681	TTTGACGCGACCAAAACCTGATATACAGAGTGGGAGAACTGTTGAAGTTATATTTGTAAGC	1740
Db	1918	CTGACGCGACCAAAACCTGATATACAGAGTGGGAGAACTGTTGAAGTTATATTTGTAAGT	1977
Oy	1741	GCTAACCCAAAGATTTACGACAGAGAACCAACCATTAACCTTTCTCACTGTGAGAGAA	1800
Db	1978	GCTAACCCGAAAGATTTACGATACAAAAACCAACCATAGACCTTCCACACTGTGAGAGAAA	2037
Oy	1801	TACGAGGACATCTGATGACACTGGCAGATATATGATTAACGATGGCTCTGGGAGACGAGG	1860
Db	2038	TATAGAGCTACTTCAACATCGTGGCAGATATGATGATATGATGCTCTCTGGAGACTGCT	2097
Oy	1861	TTTTTATTTGGCACAAGGAATATCTGGGTCTGAGCAATCAACATATATCTGGCATATTCCA	1920
Db	2098	TTTTTATTTGGCACAAGGAATCTGTGGGTCTGAGTATATCAACAGTGAAGATGGCATATTCCA	2157
Oy	1921	GATACGTGCTACCTGGATATCTACAGATAAGATATTTTGGACACAATTCGAAAGCAGAA	1980
Db	2158	GACACTGTCGCCAGCTGGAAATCTACAGATAAGATATTTTGGACACAATTCGAAAGCAGAC	2217
Oy	1981	CTTCTGAAGACCGGTGTCATACTACTGCAATTTGAAGGAATTTCTCTCTTTTGAAGTGTG	2040
Db	2218	ATTTCTGAAGCTGCTGTCTCATCTTTCATTTTGAAAGGCACTTCCCGGCTTTTGAAGTTGTA	2277
Oy	2041	ACTACTTAG 2049	
Db	2278	ACTATTTAG 2286	
RESULT 7			
BC022604		1571 bp	linear
LOCUS		mus musculus, similar to N-acetylphosphoglycine amidohydrolase 2, clone	ROD 07-AUG-2002
DEFINITION		IMAGE:4162513, mRNA, partial cds.	
ACCESSION		BC022604	
VERSION		BC022604.1	GI:18490446
KEYWORDS			
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE			
1		(bases 1 to 1571)	

**AUTHORS** Strausberg, R.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

**REMARK** NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
**COMMENT** Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mdedpaxil.stanford.edu](mailto:mdedpaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

**FEATURES** Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRM Plate: 43 Row: C Column: 22.  
 Location/Qualifiers  
 1. 1571

**source**  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="IMAGE:4162513"  
 /tissue\_type="Liver, normal. 5 month old male mouse."  
 /clone\_id="NCI.CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6"  
 <1. 1493  
 /codon\_start=3  
 /product="Similar to N-acylsphingosine amidohydrolase 2"  
 /protein\_id="AAH22604.1"  
 /db\_xref="GI:18490447"  
 /translation="KLVDINGEDLGLISFAIHPIVSMNNSHNFVSDMKVAYLFEQ  
 ENKNGYLPQGFVAFSSNIGDVSPLIGHCVNTGSCNDSTCPNGSPSCMA  
 SGPQDMFESTHIGRIIYOKAKELYASQSEVTGPVLAHQVMNTDVSOLNHT  
 VTKCPALGYSFAACTIDGVSGLNTGCTEGEDPMDLRDQLKPSSEIYECQPK  
 PHLHSGELTIHPMOPDIVDQIVTVGSLAAIIPGELTSSGRFRAIKERFLY  
 GMDKMTVIAGLSNYTHYITTYEEVOYRAASTIYGVPHLSLYIOLFRLAKAIA  
 TDIVANMSSGPEPPFKNILASLIPIRIDRAPRIKHFQVLOPAKREIVGVEYIE  
 VGANRNSHENTHOTFLVEKTEEDSVADWQIMNDASVETFRVHMKIGLSNATIT  
 WHIPTAYPGIYRIRYFGHNRKQELIKPAVILAFEGISSPEFVVT"

**CDS**  
 467 a 355 c 370 g 379 t  
 BASE COUNT  
 ORIGIN

**Query Match** 72.9% Score 1493; DB 10; Length 1571;  
**Best Local Similarity** 100.0%; Pred. No. 0;  
**Matches 1493:** Conservative 0; Mismatches 0; Indels 0; Gaps 0;

557 TGAACCTGGTGGATTGATGAGAAAGACTGGTCTTATCAGCTGGTTGCCATCCACC 616  
 1 TGAACCTGGTGGATTGATGAGAAAGACTGGTCTTATCAGCTGGTTGCCATCCACC 60  
 617 CCGTAGCATGAACATAGCAACCACTTGTGAATAGTGAATATGGCTATGGCGCTT 676  
 61 CCGTAGCATGAACATAGCAACCACTTGTGAATAGTGAATATGGCTATGGCGCTT 120  
 677 ACCTTTTGGAGCAAGAAAGAAAGAGCTATCTGCTGGACAGGAGCCGTTTGAAGCAG 736  
 121 ACCTTTTGGAGCAAGAAAGAAAGAGCTATCTGCTGGACAGGAGCCGTTTGAAGCAG 180  
 737 GCTTGGCTTCATCAATCTCGGAGAGCTGTACCCCAACATCTTGGCCCGCATTTGTCTA 796  
 181 GCTTGGCTTCATCAATCTCGGAGAGCTGTACCCCAACATCTTGGCCCGCATTTGTCTA 240  
 797 ACACAGGGAGCTCTTGTGACACAGACAGAGCACTGTCCCAAGCGTGGGCTTACATGT 856  
 241 ACACAGGGAGCTCTTGTGACACAGACAGAGCACTGTCCCAAGCGTGGGCTTACATGT 300

857 GCATGCCACGCGGACCTGGACAAAGACATTTTGAGAGCACACACATTTATAGACGATCA 916  
 301 GCATGCCACGCGGACCTGGACAAAGACATTTTGAGAGCACACACATTTATAGACGATCA 360  
 917 TCTATCAGAAAGGCAAGAGAGCTGTATGCTTGGCTCCCGAGAGAGGAGACCGGCCAGTGC 976  
 361 TCTATCAGAAAGGCAAGAGAGCTGTATGCTTGGCTCCCGAGAGAGTGAACCGGCCAGTGC 420  
 977 TTGCAGCTCAGCAGTGGGTGAACATGACAGATGTGAGCTGCAGCTCAATGGCCACACACA 1036  
 421 TTGCAGCTCAGCAGTGGGTGAACATGACAGATGTGAGCTGCAGCTCAATGGCCACACACA 480  
 1037 CAGTGAAGAGCTGTAAACCTGCGCTGGCTTACAGTTTGGCCGAGGACCAATTTGATGGAG 1096  
 481 CAGTGAAGAGCTGTAAACCTGCGCTGGCTTACAGTTTGGCCGAGGACCAATTTGATGGAG 540  
 1097 TTTCGGGCTCAATATTTACACAGGAAACACAGAGAGGAGATCCATTTGCGACACTCTTC 1156  
 541 TTTCGGGCTCAATATTTACACAGGAAACACAGAGAGGAGATCCATTTGCGACACTCTTC 600  
 1157 GGGACCAAGCTCTTGGGAAACCAATCTGAAGAGATTTAGAGTGTAGAAAACCAACCA 1216  
 601 GGGACCAAGCTCTTGGGAAACCAATCTGAAGAGATTTAGAGTGTAGAAAACCAACCA 660  
 1217 TCCTGCTTACAGTGGAGAGCTGACGATACCAATCCTTGGCAACCAATTTGTTGATG 1276  
 661 TCCTGCTTACAGTGGAGAGCTGACGATACCAATCCTTGGCAACCAATTTGTTGATG 720  
 1277 TTCAGATTTTACCGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 1336  
 721 TTCAGATTTTACCGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTT 780  
 1337 TGTGGGAGCAAGATTTTGTGAGCAATTTAAAAAAGATTTTACATTTATGGAGTGAAG 1396  
 781 TGTGGGAGCAAGATTTTGTGAGCAATTTAAAAAAGATTTTACATTTATGGAGTGAAG 840  
 1397 ATATGACCTTGTATTCGAGAGCTTACAGATTAATTTATACATTTATACATATATG 1456  
 841 ATATGACCTTGTATTCGAGAGCTTACAGATTAATTTATACATTTATACATATATG 900  
 1457 AAGAATACAGAGCTGACGCGTACGAGGAGCATCTTACATATATGAGGACACACACCTGT 1516  
 901 AAGAATACAGAGCTGACGCGTACGAGGAGCATCTTACATATATGAGGACACACACCTGT 960  
 1517 CTGCATACATCCAACTCTTACAGAGCTTGTGAAGCAATTTCTACGACACAGTACCA 1576  
 961 CTGCATACATCCAACTCTTACAGAGCTTGTGAAGCAATTTCTACGACACAGTACCA 1020  
 1577 ACATGAGAGTGGTCCCGAGGCTCCATCTTCAAAAATCTATACCTTCACTTATTCCTA 1636  
 1021 ACATGAGAGTGGTCCCGAGGCTCCATCTTCAAAAATCTATACCTTCACTTATTCCTA 1080  
 1637 ATATTGGCGATGAGACCAATTTGGCAAAACATTTTGGGAGTCTTGGACCGCAACAAAC 1696  
 1081 ATATTGGCGATGAGACCAATTTGGCAAAACATTTTGGGAGTCTTGGACCGCAACAAAC 1140  
 1697 CTGAATACAGAGTGGAGAGTGTGAAGTTATTTGTAGGCGCTAACCAAGAAAT 1756  
 1141 CTGAATACAGAGTGGAGAGTGTGAAGTTATTTGTAGGCGCTAACCAAGAAAT 1200  
 1757 CAGCAGAGAGACAGACCCATCAAACTTCCACGCTGAGAGAAATACAGAGCTCTGTAG 1816  
 1201 CAGCAGAGAGACAGACCCATCAAACTTCCACGCTGAGAGAAATACAGAGCTCTGTAG 1260  
 1817 CTGACTGGCAGATTAATGTATAAGATGCTCTCTGGAGAGAGGTTTATTTGGCAAGAG 1876  
 1261 CTGACTGGCAGATTAATGTATAAGATGCTCTCTGGAGAGAGGTTTATTTGGCAAGAG 1320  
 1877 GAATACCTGGGCTGTGAGCAATGCAACATATACCTGGCATATTCAGATACCTGCTACCTG 1936  
 1321 GAATACCTGGGCTGTGAGCAATGCAACATATACCTGGCATATTCAGATACCTGCTACCTG 1380  
 1937 GAATCTACAGAAATTAAGATTTTGTGAGACACATTCGGAAGCAGAGACTTGTGAACCCGCTG 1996



Db 321 ACTTATATACATGATATGATATTAATTAACGGGTACACATATTCATTCAGTCTCTGT 380  
 Oy 307 GGGTTTTCATATATACACTCTATATACCGCCAGCAGAGATAGCAACCGGACCTTT 366  
 Db 381 GGTTTTACAGATATGATATATATGATATACAGATTAAGTTTCTATATTAAGAATTTTC 440  
 Oy 367 CAGTACATATGCTCTGGATCATGAAGACATTTGATATAGCTCACACAATCTTAACCA 426  
 Db 441 GATACAAATTTGATGATATGCTTCACGAATTTGTAAGATTCATTAATTCGGTCAACCT 500  
 Oy 427 GGCAAAATCTTATTAACAAGGAATGTGCTATATGTCAGATATCAACCGAACCCCTTC 486  
 Db 501 GCAGAAATGTTAACTCAACAGAGTGAATTAAGAAATTCATAATTCGTTCCACCAAT 560  
 Oy 487 TCTTACCTTCGATCCACAGTACAGAGAGAGAGATTTCTTCAACACAGACAGAGAA 546  
 Db 561 GCATATGATATATATCCAGAGAGAGAGAGAGAGATATGATATGATATGATATGAAT 620  
 Oy 547 ATGCTGCTTGAACCTGGTGAATTTGAATGAGAGAAAGCTTGGTCTTATACAGTGT 606  
 Db 621 ATGACAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 680  
 Oy 607 GCATACCAACCCGTAGATGATACATAGCAACCACTTGTGATATGATGATGATGATGAT 666  
 Db 681 GGTGTTCAATGATCTTCAATGATATATCAATTAATCTTGTGATATGATGATGAT 740  
 Oy 667 TATGCGCTTACCTTTTGGT---AGCAAGAAAGAAAGAGCTATCTCTGAGACAGGA 723  
 Db 741 TATGCAATGATATCTTTGGGAAAGAAAGATGCATGAGCAAGAGCTATTCACAGTACT 800  
 Oy 724 CCGTTTGTAGCAGCTTTGCTTATCAATCTCGAGAGAGCTGACCCCAATCTTGGC 783  
 Db 801 CCATTCATTTGACAGCTTTGCTCAAGATGATGATGATGATGATGATGATGATGATGAT 860  
 Oy 784 CCGCATGCTGTCACACAGAGGAGTCTGTGACACAGCAAGAGACCTGTCACCAAGGT 843  
 Db 861 CCACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 920  
 Oy 844 GGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 903  
 Db 921 GAGAGGTGTTGGGCACTAGGACAGGATGATGATGATGATGATGATGATGATGATGAT 980  
 Oy 904 ATAGAGCGATCATATATCAGAGAGCCCAAGAGCTGTATGCTCTGCTCCAGAGAGTG 963  
 Db 981 ATTTGTTGATATCAATTTCAACAGGATGATGATGATGATGATGATGATGATGATGAT 1040  
 Oy 964 ACCGCCAGTGTGACAGCTCACCAGTGGGAGAACATGACAGATGAGAGCTGTC--- 1018  
 Db 1041 AGTGTAGATTTCAATATAGACACACATGAGAAACCAATTTACCAATGTTCACTTAAGCA 1100  
 Oy 1019 ---AGCTCAATGCCACACACACAGTGAAGAGCTGAACCTGGCTGGCTACAGTTT 1074  
 Db 1101 CCATATATATAGTGGGTTGAGGGGCCAACACATGATAGAGGTGCAATGGTATAGTTT 1160  
 Oy 1075 GCCGAGCAGCAATGATGAGATTTGGGCTCATATTTACACAGGAGACTACGAGAGG 1134  
 Db 1161 GCTGTGTGATCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1220  
 Oy 1135 GATCATCTGTGGACACTCTTGGGACAGCTCTTGGGAAACCAATCTGAAGAGATTTGA 1194  
 Db 1221 ACTATATCATTTTGAATTTTCACTGCTGTGATCATTTGCAAAACCAACACACACAAACA 1280  
 Oy 1195 GAGTGTCAAGAAACCAACCAATCTGCTTCACTGAGAGAGCTACGATACCATCT 1254  
 Db 1281 GCTTGTCAAGCTCCAAACCAATTTTAATGTA---TGTTGTATGTTGAACCAATCCCA 1337  
 Oy 1255 TGGCAGCAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1314  
 Db 1338 TGGGTATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1397  
 Oy 1315 ATCCCTGGGATTTAAACCACTGTCGGAGACAGATTTCTGAGGCAATTTAAAAAGAA 1374

Db 1398 GTTCCAGTGAATTCACAAACAATGCTGTGATAGAAATTAAGAAATACCGTTAGAGAAATC 1457  
 Oy 1375 TTTCACATTTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434  
 Db 1458 ATTTG---GTCAATCAATTTGAATTTGAATTTGATTTATTTGCTGTTAGCAATCTTAT 1514  
 Oy 1435 ACATATTAATTTACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1494  
 Db 1515 TCAGTATATTTGCAACATTTGAGAGATTTGCAAGTTCAAAGATATGAGAGTGCATCAAC 1574  
 Oy 1495 ATCTATGACACACACACCTGCTGCTGATATATCCAACTTCAGAGCTGCTTAAGCA 1554  
 Db 1575 GTTTTGGACCTCATCTTTAGAGCATATTCACAGAAATTTGCAAACTTTGACACATCA 1634  
 Oy 1555 ATTTGCTAGGACACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1614  
 Db 1635 ATTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1694  
 Oy 1615 CTAATAGCTTCACTTATTTCTATATTTGCGGATGATGATGATGATGATGATGATGATGAT 1674  
 Db 1695 CCATCTCTTTTAC---CACAGATCATTTGTTGATGATGATGATGATGATGATGATGATGAT 1751  
 Oy 1675 GATGCTTGGACCCGACCAAAACCGATATGATGATGATGATGATGATGATGATGATGAT 1734  
 Db 1752 GATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1811  
 Oy 1735 GTAGCGCTAACCCAAAGATTTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794  
 Db 1812 ATTTCTATGTTGGCCAACTTTGAATTAATTTATGATGATGATGATGATGATGATGATGAT 1871  
 Oy 1795 GAGAAATACGAGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1854  
 Db 1872 GATCAAT---TACATCTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1928  
 Oy 1855 ACAGAGTTTATTTGACCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914  
 Db 1929 ACMAATTTATTTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1988  
 Oy 1915 ATTCAGATTTCTGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1973  
 Db 1989 ATCTCCAAATCATCTCAACAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2047

RESULT 10  
 AC117816 165407 bp DNA linear HTG 16-AUG-2002  
 LOCUS Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE, 12  
 DEFINITION unordered pieces.  
 AC117816.2 GI:22267711  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 165407)  
 AUTHORS Birren, B., Nusbaum, C. and Lander, E.  
 TITLE Mus musculus, clone RP24-573C13  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 165407)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barre, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
 Boukhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collumore, A.,  
 Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
 Fero, S., Ferreira, P., FitzHugh, M., Gage, D., Galagan, J., Gardyna, S.,  
 Glode, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Lakoque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, R., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meidrid, J., Menees, L.,  
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,



## TITLE

## JOURNAL

REFERENCE  
AUTHORS

Notbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vasiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Submitted (10-Apr-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
(bases 1 to 165407)

Barren,B., Nusbam,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barron,N., Bastien,V., Bloom,T., Boguslavsky,L., Buchhalter,B.,  
Cammarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diz,T.S., Dodge,S.,  
Fato,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mjengwa,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Notbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Testaye,S., Theodore,J., Topham,K., Travers,M., Vasiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## COMMENT

Submitted (16-Aug-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 16, 2002 this sequence version replaced gti:20128606.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: 125854

Center clone name: 573.C13

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 160616 bases at least Q40

Consensus quality: 162867 bases at least Q30

Consensus quality: 163827 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 164307; sum-of-coverage

Quality coverage: 7.1 in Q20 bases; agarose-fp

Quality coverage: 7.2 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 705: contig of 705 bp in length  
\* 706 805: gap of 100 bp  
\* 806 1789: contig of 984 bp in length  
\* 1790 1889: gap of 100 bp  
\* 1890 3396: contig of 1507 bp in length  
\* 3397 3496: gap of 100 bp  
\* 3497 4739: contig of 1243 bp in length

\* 4740 4839: gap of 100 bp  
\* 4840 11073: contig of 6234 bp in length  
\* 11074 11173: gap of 100 bp  
\* 11174 18734: contig of 7561 bp in length  
\* 18735 18834: gap of 100 bp  
\* 18835 29191: contig of 10357 bp in length  
\* 29192 29291: gap of 100 bp  
\* 29292 42552: contig of 1361 bp in length  
\* 42553 42652: gap of 100 bp  
\* 42653 60018: contig of 17365 bp in length  
\* 60019 60118: gap of 100 bp  
\* 60119 81529: contig of 21411 bp in length  
\* 81530 81629: gap of 100 bp  
\* 81630 120500: contig of 38871 bp in length  
\* 120501 120600: gap of 100 bp  
\* 120601 165407: contig of 44807 bp in length.

## FEATURES

## source

1. 165407

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="RP24-573C13"

/clone\_11b="RP24 Male Mouse BAC"

1. 705

/note="assembly-fragment"

806. 1789

/note="assembly-fragment"

1890. 3396

/note="assembly-fragment"

3497. 4739

/note="assembly-fragment"

4840. 11073

/note="assembly-fragment"

11174. 18734

/note="assembly-fragment"

18835. 29191

/note="assembly-fragment"

29292. 42552

/note="assembly-fragment"

42653. 60018

/note="assembly-fragment"

60119. 81529

/note="assembly-fragment"

81630. 120500

/note="assembly-fragment"

120601. 165407

/note="assembly-fragment"

misc\_count 46319 a 32936 c 33090 g 51958 t 1104 others

## BASE COUNT

ORIGIN

Query Match 9.4%; Score 192.4; DB 2; Length 165407;

Best Local Similarity 92.7%; Pred. No. 26-46; 16; Indels 0; Gaps 0;

Matches 202; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

1832 TGTATAGAGATGCTCTGCGAGAGAGGTTTATTTGACACAAAGAAATGAGTCTGA 1891

1427 TTTATTTATTTATTTGTTTCTATTACAGGTTTATTTGACAAAGAAATGAGTCTGA 1486

1892 GCAATGCAACATATATCTGCGATTTCCAGATATCTGCTACCTGGAATCTACAGATTA 1951

1487 GCAATGCAACATATATCTGCGATTTCCAGATATCTGCTACCTGGAATCTACAGATTA 1546

1952 GATATTTTGGACACAAATGGAAGGAGGAACTTGAACCCGCTGCACTACTAGCATTTG 2011

1547 GATATTTTGGACACAAATGGAAGGAGGAACTTGAACCCGCTGCACTACTAGCATTTG 1606

2012 AAGGAATTTCTCTCTCTTTTGAAGTTGTCACTACTTAG 2049

1607 AAGGAATTTCTCTCTCTTTTGAAGTTGTCACTACTTAG 1644

RESULT 11

AC117816/c

LOCUS AC117816 165407 bp DNA linear HTG 16-AUG-2002

```

DEFINITION Mus musculus clone RP24-573C13, WORKING DRAFT SEQUENCE. 12
unordered pieces.
ACCESSION AC117816
VERSION AC117816.2 GI:22267711
KEYWORDS HTGS_PHASEL, HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 165407)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
JOURNAL Unpublished
TITLE 2 (bases 1 to 165407)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouckghalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Labrecque, K., Lamazares, R.,
Landers, T., Lehoczký, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Menus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Neill, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, P., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Struass, K., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zaloun, J., Zembek, L., Zimmer, A. and Zody, M.
JOURNAL Direct Submission
TITLE Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 165407)
REFERENCE Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckghalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Neill, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zaloun, J.,
Zembek, L., Zimmer, A. and Zody, M.
JOURNAL Direct Submission
TITLE Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 16, 2002 this sequence version replaced g1:20128606.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996.1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L25854
Center clone name: 573_C_13

```

```

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160616 bases at least Q40
Consensus quality: 162967 bases at least Q30
Consensus quality: 163827 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 164307; sum-of-ctrls
Quality coverage: 7.1 in Q20 bases; agarose-fp
Quality coverage: 7.2 in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 12 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 705: contig of 705 bp in length
706 805: gap of 100 bp
806 1789: contig of 984 bp in length
1790 1889: gap of 100 bp
1890 3396: contig of 1507 bp in length
3397 3496: gap of 100 bp
3497 4739: contig of 1243 bp in length
4740 4839: gap of 100 bp
4840 11073: contig of 6233 bp in length
11074 11173: gap of 100 bp
11174 18734: contig of 7561 bp in length
18735 18834: gap of 100 bp
18835 29191: contig of 10337 bp in length
29192 29291: gap of 100 bp
29292 42552: contig of 13261 bp in length
42553 42652: gap of 100 bp
42653 60018: contig of 17366 bp in length
60019 60118: gap of 100 bp
60119 81529: contig of 21411 bp in length
81530 81629: gap of 100 bp
81630 120500: contig of 38871 bp in length
120501 120601: gap of 100 bp
120601 165407: contig of 44807 bp in length.
-----
FEATURES
source
1. 165407
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-573C13"
/clone_lib="RPCT-24 Male Mouse BAC"
1. 705
/note="assembly-fragment"
806. 1789
/note="assembly-fragment"
1890. 3396
/note="assembly-fragment"
3497. 4739
/note="assembly-fragment"
4840. 11073
/note="assembly-fragment"
11174. 18734
/note="assembly-fragment"
18835. 29191
/note="assembly-fragment"
29292. 42552
/note="assembly-fragment"
42653. 60018
/note="assembly-fragment"
60119. 81529
/note="assembly-fragment"
81630. 120500
/note="assembly-fragment"
120601. 165407
/note="assembly-fragment"

```





\* 30962 31061: gap of 100 bp  
\* 31062 41639: contig of 10578 bp in length  
\* 41640 41739: gap of 100 bp  
\* 41740 77584: contig of 35845 bp in length  
\* 77585 77684: gap of 100 bp  
\* 77685 107166: contig of 29482 bp in length  
\* 107167 107266: gap of 100 bp  
\* 107267 140582: contig of 33316 bp in length  
\* 140583 140682: gap of 100 bp  
\* 140683 174914: contig of 34232 bp in length  
\* 174915 175014: gap of 100 bp  
\* 175015 184101: contig of 9087 bp in length.  
Location/Qualifiers

## FEATURES

source

1. 184101  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-346D12"  
/clone\_lib="RPC1-23 Female Mouse BMC"  
1. 1186  
/note="assembly-fragment  
clone\_end:SP6  
vector\_side:left"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

misc\_feature

/note="assembly-fragment"

BASE COUNT 51100 a 39204 c 38438 g 53758 t 1601 others  
ORIGIN

Query Match 8.8%; Score 179.8; DB 2: Length 184101;  
Best Local Similarity 98.9%; Pred. No. 1.4e-42;  
Matches 181; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 211 TTGAGTCTCGAAGAGAGAGAGTAAATGCTGCTCTGTATCGAAGAGACAATGTT 270  
DB 17401 TTCAAGGTCTCTGAGAGAGAGAGTAAATGCTGCTCTGTATCGAAGAGACAATGTT 17460  
OY 271 ATCTGAGTGCATTCACACACACTCTGCCCGAGGAGGTTTTCATATACACTCTAT 330  
DB 17461 ATCTGAGTGCATTCACACACACTCTGCCCGAGGAGGTTTTCATATACACTCTAT 17520  
OY 331 ATACTGCCGAGGAGGATTCAGCAACGGAGCTTTCAGTACATAGTCTGCGATCATG 390  
|||||

DB 17521 ATACTGCCGAGGAGGATTCAGCAACGGAGCTTTCAGTACATAGTCTGCGATCATG 17580  
OY 391 AAG 393  
DB 17581 AAG 17583  
|||

RESULT 13  
HUMD4C12M5 235 bp mRNA linear PRI 04-FEB-1999  
LOCUS  
DEFINITION Human HepG3 partial cDNA, clone hmd4c12m5.  
ACCESSION D17062.1 GI:598818  
VERSION  
KEYWORDS gene signature.  
SOURCE Homo sapiens Male cell\_line:HepG2 cDNA to mRNA, clone\_lib:Kiserv.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 235)  
TITLE Matoba,R., Okubo,K., Hori,N., Fukushima,A. and Matsubara,K.  
The addition of 5'-coding information to a 3'-directed cDNA library  
Improves analysis of gene expression  
JOURNAL Gene 146 (2), 199-207 (1994)  
MEDLINE 94357437  
REFERENCE 2 (bases 1 to 235)  
AUTHORS Matoba,R.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUL-1993) Ryo Matoba, Osaka University, Institute for  
Molecular and Cellular Bio: 1-3, Yamada-oka, Suita, Osaka 565,  
Japan (E-mail:matoba@imh.ribo.osaka-u.ac.jp,  
Tel:81-6-877-5111(ex.3314), Fax:81-6-877-1922)  
Submitted (21-Jul-1993) to DDBJ by:  
Ryo Matoba  
Molecular Microbiology and Genetics Lab.  
Research Institute of Innovative Technology for the Earth 9-2  
Kizugawadai Kizu-cyo,  
Soraku-gun, Kyoto  
Japan, 619-02  
Phone: 07747-5-2308  
Fax: 07747-5-2321.

## FEATURES

source

## COMMENT

1. 235  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="Male"  
/cell\_line="HepG2"  
/clone\_lib="Kiserv"  
Location/Qualifiers

BASE COUNT 64 a 46 c 57 g 68 t  
ORIGIN

Query Match 7.7%; Score 158; DB 9; Length 235;  
Best Local Similarity 85.8%; Pred. No. 2.4e-36;  
Matches 187; Conservative 0; Mismatches 30; Indels 1; Gaps 1;

OY 1832 TGTATACGATGCTCTGCGAGAGAGGTTTATTTGGCACAAGCAATACCTGCGA 1891  
DB 2 TGTATACGATGCTCTGCGAGAGGTTTATTTGGCACAAGCAATACCTGCGGT-TGA 60  
OY 1892 GCAATGCAACATATACGATGCTCTGCGAGAGGTTTATTTGGCACAAGCAATACCTGCGA 1951  
DB 61 GCAATGCAACATATACGATGCTCTGCGAGAGGTTTATTTGGCACAAGCAATACCTGCGA 120  
OY 1952 GATATTTGGACACATGCGAAGCAAGCAACTCTGAAACCCGCTGCTACTAGCATTTG 2011  
DB 121 GATATTTGGACACATGCGAAGCAAGCAACTCTGAAACCCGCTGCTACTAGCATTTG 180  
OY 2012 AAGCAATTTCTTCCTTTTGAAGTTGCATCTATTAG 2049  
DB 181 AAGCAATTTCTTCCTTTTGAAGTTGCATCTATTAG 218  
RESULT 14  
AC012788

LOCUS AC012788 53059 bp DNA linear HTG 03-NOV-1999  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
 pieces.  
 ACCESSION AC012788  
 VERSION AC012788.1 GI:6223530  
 KEYWORDS HTG; PHASE2.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 53059)  
 AUTHORS Adams, M. and Venter, J. C.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT This sequence was identified as CDM:10210280 by the submitter.  
 For further information on this sequence you may e-mail to  
 fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 FEATURES  
 source  
 1..53059  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 BASE COUNT 14592 a 11833 c 12043 g 14591 t  
 ORIGIN  
 Query Match 7.6%; Score 155.2; DB 2; Length 53059;  
 Best Local Similarity 50.6%; Pred. No. 3.7e-35;  
 Matches 640; Conservative 0; Mismatches 528; Indels 98; Gaps 7;  
 QY 64 TTGATGGGCTATGGCAAAATGGCCAGATATCAGGGGCTCTCTCAGCAGCTGTTCAGC 123  
 Db 35557 TAGATGGGCTATGGCAAAATGGCCAGATATCAGGGGCTCTCTCAGCAGCTGTTCAGC 35616  
 QY 124 CGTCTCTTTATCTGGGGGATCCAGATGGGCTCAATGCAATGGCATTTGTAGGGGTGAA 183  
 Db 35617 CGTCTCTTTGTGGTGGAGATGAAAAGGC---AACGAGTGGCTTGTGAGCGCAAT 35673  
 QY 184 CTATGATATTTTCCCAAGACTGAGTTGAGGCTCTGAAAGACTAGAGATTAATAT 243  
 Db 35674 GCCGAATGATGGCTACGATTAAGAGGAGGATGATAAAGCACTCCAGCACTGAT 35733  
 QY 244 GGCTCTCTGATCCAGAGCAATGTTATCTGATGCTCATTCACAGCACTTGGCCCA 303  
 Db 35734 GGCAACATCTACCAATGATTAATGTGGCATCAGCCACACGCGTCTCTCT 35793  
 QY 304 GCAGGTTTTCCTCAATATACCTATATATCTCGCCAGCGAGGATTCAGCAACCGGAC 363  
 Db 35794 GCGCGATTCCTGATGATCTCTCTATGACATCTTCATTTCTGGGCTTGTGCTAGACG 35853  
 QY 364 TTTGAGTACATAGTCTCTG-----GATCATGAAG 393  
 Db 35854 TTTGAGGATGGCTCAGGAGCTGATCTGTGGGTTTTCATTTCAACTGAATTTGTAC 35913  
 QY 394 AGCATGATATAGCTCACACAATCTTAACACAGCAAA-----GATCATGAAG 432  
 Db 35914 GGAACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 35973  
 QY 433 -----ATCTTATCAACAAAGAAATGTTGTAATGTGCTAGATCAAGCAAGCCCT 484  
 Db 35974 ATGATGCGATCTTCTGCTGCAAACTACTGTGCTTAATGTTAATCAATCAATGTTGCT 36033  
 QY 485 CCTCTTACCTTCTGATTCACAGTCCAGAGAGCAAGATATTTCTTAAACACAGCAAGG 544  
 Db 36034 CATCTTACTTGAAGATCCGCCAGAGAGCTGCCAATATAGAGAGATGAGATGAGTAAAG 36093  
 QY 545 AAATGCTGGCTTAAACTGATGATTTGAATGAGAAAGACTTGGGTTTATTCAGCTGGT 604  
 Db 36094 CACTGACCCAGCTGATTTGTGACCTGGAAGAAACCACTCTGAGGCTTTCAACATGGT 36153

QY 605 TTGCCATCCACCCGTGAGCATGAACATAGCAACCACTTTGTGATATGACATATAG 664  
 Db 36154 ATGCGGTCAATGCCACTCTCCATGATACCAATACAGACTGTGACAGCAATGTGG 36213  
 QY 665 GCTATGCGGCTTACCTTTTGTAGCAAGAAAGAACAAAGGCTATCTGCTGAGAGGAC 724  
 Db 36214 GTTATGCGGCTCTGCTGTAAGAAAGATGCAATCCGAAAGATGCCGGAAGGCA 36273  
 QY 725 CGTTTGTAGCAGGCTTCTCTCATCAATCTCGAGAGCTGTACCCCAATTTTGGCC 784  
 Db 36274 AGTTGCTGGTGGCTCTTCTGCTATCAACCTTGGGATGTCTCCCAATATATGGTCT 36333  
 QY 785 CGCATTTGTCAACACAGGGGAGTCTGTGCAACGACAGCAAGCAACCTGCCAAGCTG 844  
 Db 36334 CCAGTGTCTCATCTCCGCAATGATGTGATCTACTGCTCCGCTTGTCCATGCGC 36393  
 QY 845 GCGCTAGCATGTGATGCGCAGCGGACCTGAGCAAGACATGTTGAGAGCACACATTA 904  
 Db 36394 AG---GGAGATTTGCTTCTCTCCGAGCCCGTAAGGATATGTTGAAAGCACACCATCT 36450  
 QY 905 TAGACGATATCTATCAGAAAGCCAAAGAGCTG-----TATGCTCTG 949  
 Db 36451 TGGGTCAACGTCTGGCGATCTCTCTGGGACTCTCAACGAAACAGACAGAGTCCA 36510  
 QY 950 CCTCCAGAGAGTGAACCGCCAGTCTTGGAGCTCAACAGTGGGTGAACATGACAGATG 1009  
 Db 36511 CGGACGGGAGGCTCACTGGAATGTAGAGTTCACTACCAAGTTGTGACATGCCCACT 36570  
 QY 1010 TGAGGCTC-----CAGCTCAATGGCCACACACACAGTGAAGAGCTGAAC 1054  
 Db 36571 ACAATGGCAGACCTACAAATCCCTGACACAGAGAGGTGACAAAGATCAGGGGATGTACG 36630  
 QY 1055 CTGCGCTGGCTTACAGTTTTCGCGAGCAACATTTGATGAGTTTCGGGCTCAATATTA 1114  
 Db 36631 CGGCATGGGCTTAATAGCTTCTCTGTGTACACAGAGTCACTGAGCTTCAAGCTTCG 36690  
 QY 1115 CACAGGAATACGGAAGGAGTCAATCTGGGACACTCTCGGAGCAGCTTGGGAA 1174  
 Db 36691 AGCAGGAAACCAACAGCAGCAACCCATGTGMACTTTGTGCGGAC---TTCAATCGTG 36747  
 QY 1175 AACCATGTAAGATTTAGAGTGTCAAAAACCAACCAATCTGCTTCAAGATGAG 1234  
 Db 36748 CTCACAGCAGAGAGATCAATGAAGTCCATGAGCCCAACCCATTTCTGCCACAGCA 36807  
 QY 1235 AGCTGA 1240  
 Db 36808 GGGTGA 36813  
 RESULT 15  
 AC008299  
 LOCUS AC008299 173844 bp DNA linear INV 14-FEB-2001  
 DEFINITION Drosophila melanogaster, chromosome 3R, region 99F1-99F11, BAC  
 clone BACR48G16, complete sequence.  
 ACCESSION AC008299  
 VERSION AC008299.5 GI:12830065  
 KEYWORDS HTG.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 173844)  
 AUTHORS Celisier, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,  
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amaratilake, P.G., Brandon, R.C.,  
 Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,  
 Carlson, J.W., Center, A., Chemp, M., Davenport, L.B., Dietz, S.M.,  
 Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,  
 Ferreira, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A.,  
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, R.J.,  
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,  
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Pachleu, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
Phouanavong, S., Pitman, G. S., Puri, V., Richards, S., Scheeler, F.,  
Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S. M.,  
Zaveri, J. S., Smith, H. O., Rubin, G. M., and Venter, D. C.  
Sequencing of *Drosophila* chromosome 3R, region 99F4-99F11  
Unpublished  
2 (bases 1 to 173844)

TITLE  
JOURNAL

Submitted (02-AUG-1999) *Drosophila* Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

## COMMENT

On Feb 14, 2001 this sequence version replaced g1:6633925.  
Sequence submitted by:  
Berkeley *Drosophila* Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).  
Location/Qualifiers

FEATURES  
source

1..173844  
/organism="Drosophila melanogaster"  
/strain="Y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
/map="99F4-99F11"  
/clone="BAC BACR48G16 (D559)"  
/clone\_11b="RPL-98 (Ruswell) Park Cancer Institute  
*Drosophila* melanogaster BAC library, partial EcoRI in  
pBAC3.6)"  
BASE COUNT 48763 a 38515 c 37719 g 48847 t

Query Match 7.6%; Score 155.2; DB 3; Length 173844;  
Best Local Similarity 50.6%; Freq. No. 4.4e-35;  
Matches 640; Conservative 0; Mismatches 528; Indels 98; Gaps 7;

QY 64 TTGATGGCTATGCGCAAAATGCGCAATGCGGGGTCTCTCCACGAGCTGTTCAGC 123  
Db 33798 TGATGGCTATGCGCAAAATGCGCAATGCGGGGTCTCTCCACGAGCTGTTCAGC 33857  
QY 124 CGTCTTTTATCTTGGCGGATCCAGATGGGTCAATTCGAATGGCATTTGTGACCGTGA 183  
Db 33858 CGTCTTTTATCTTGGCGGATCCAGATGGGTCAATTCGAATGGCATTTGTGACCGTGA 33914  
QY 184 CTATGATGATTTCCCAAGACTGAGGTGAGGTGCTGGAAGACTAGAGCTAAATAT 243  
Db 33915 GCGGATATATGGCTATGAGATTTGAGAGGAGGTATATAAAGACTCCAGGACGTTAT 33974  
QY 244 GGCCTCTCTATCGAAGACAATGTTATCTGAGTGCATTCACACACTCTGGCCCA 303  
Db 33975 GGCACATCTACACAAATGATATGCGCATGAGGACCCACACAGCGGTGCTCT 34034  
QY 304 GCAGGCTTTTCCAAATATACACTCTATATATCTGCCAGGAGGATTCAGCAACGAGAC 363  
Db 34035 GCGGATTTCTCTGATGATCTGCTATGACATCTCCATCTGGGCTTTGTGCTCCTCA 34094  
QY 364 TTTCAGTACATAGTCTCTGG-----GATCATGAAG 393  
Db 34095 TTTCAGTACATAGTCTCTGG-----GATCATGAAG 34154  
QY 394 AGCATTTGATATGCTTCACAAATCTTAAACGAGCAAA-----432  
Db 34155 GGAAGCAATATACATATTTATATATATAGTGCATCAAAAGGCGGACGACACACTGTGG 34214

QY 433 -----ATCTTTATCAACAAAGAAATGCTCTAATGTGCATCAACCGAGCCCT 484  
Db 34215 ATGTCGATCTTCTGCTCCAAACATCTGCTAATGTATACATCAATCTGCTCCCT 34274  
QY 485 CCTTACCTTCTGATATCCACACTGAGAGAGAGATTTCTTCAACACAGCAAG 544  
Db 34275 CATCTTACTAGAGAAATCCCGGAGAGAGGTGCCCAATTCAGCACAGATGATTA 34334  
QY 545 AAATGCTGTGCTTGAACACTGATGATTTGATGAGAGACCTTGGGCTTATCAGCTGT 604  
Db 34335 CACTGACCCAGCTGCGATTTGTGACCTGGAACAAACCTCCCGGGCGCTTCAACTGG 34394  
QY 605 TTGCAATCCACCCCGTGAAGATTAACATTCAGACACTTGTGAATAGTACATATG 664  
Db 34395 ATCGGCTGATGCTCCATCCATGACAAATCCAAACAGACTGAGACAGCAAGTGG 34454  
QY 665 GCTATGCGGCTTACCTTTTGGACCAAGAAAGAAAGAGCTTCTGCTGAGAGGAC 724  
Db 34455 GTTATGCGGCTTCTGCTGGAAGAGATGACAAATCCGAACAGATGCGGAAAGGCA 34514  
QY 725 CGTTTGTAGAGGCTTGTGCTTCAATCAATCTCGAGAGCTGTCACCCCACTTGTGGC 784  
Db 34515 AGTTGCTGGGTGCTTCTGCTCACTCAACCTTGGCATGTGCTCCCAATATATGGTTC 34574  
QY 785 CGCATTTGTCAACACAGGAGTCTTGTGACACAGACAGACCTGCTCCACAGGTG 844  
Db 34575 CCAAGTGTCTGATCTCCGCAATGATGATCTGACATCCGCTTGTCCACATGGG 34634  
QY 845 GGCCTGACATGTCATGCGCAGGAGGACCTGCAACAAACATGTTGAGACACACATTA 904  
Db 34635 AG-----GAGATTTCTTCTGCTCCGAGCCGCTGAGATATGTTGAAAGACCCACATCT 34691  
QY 905 TAGAGGATCATCTATCAAGAAAGCCAGAGCTG-----TATGCTCTG 949  
Db 34692 TGAGTCAAGCTCTGCGGAGATGCTGCTGTGAGATGCTCAACGAACAGAGCAAGATTC 34751  
QY 950 CTTCCAGAGAGTGAACCGGCCAGTCTTGCAGCTCAACAGTGGTGAACATGACAGATG 1009  
Db 34752 CGGCACGAGGAGTCACTGAGATGATGAGTTTCACACAGTGTGTGGACATCCCACT 34811  
QY 1010 TGAGCGTC-----CAGCTCAATGCCACACACAGTGAAGAGCTGTAAC 1054  
Db 34812 ACATGAGCAGACCTTCAATCCGCTGAGCAGAGGTCGACAAAGATCAGGGATGTCAGC 34871  
QY 1055 CTGCTCTGAGCTAGTTTGTGCGCAGGACAAATGATGAGATTTGGGCTCAATATTA 1114  
Db 34872 CGGCATGAGGCTATGATTTGTGCTGTGATCCAGCATGAGCTGAGGCTTCAAGCTTCG 34931  
QY 1115 CACAGGAACTACGGAAGGGGATTCATTTCTGGACACTCTGCGGACAGCTTTGGGA 1174  
Db 34932 AGCAGGAAACACACAGGACAAACCCATGGAACCTTTGTGCGGAC-----TTATGCTG 34988  
QY 1175 AACCTGTAAGATTTAGATGTCAGAAACCAACCAATCTGCTTCAACAGTGAAG 1234  
Db 34989 CTCACAGCAGAGGACATCAATGATGCTGAGGCCAACGCCATTTCTTGTGCGCACGGCA 35048  
QY 1235 AGCTGA 1240  
Db 35049 GGGTGA 35054

Search completed: July 5, 2003, 11:30:33  
Job time : 5399 secs





Db 210 AYLT-----DPAEDA-----LGLPSGGRYDI-----PLVLSKYYNADGTLKTSVG 252  
QY 425 DVQIVTVGSLAIAIPELITMSGRFR-----EAIKKEFALYMK-DMYV-----VIA 472  
Db 253 EDSWSWGDIIHVGQCPFLNVEPRKRLFLNAAVSRNALYFVKODMNTRLPFOVIA 312  
QY 473 GLSNVYHYITTYEY--QAQRYEASTI--YGPHTLSAVITQLFRDLAKA-----518  
Db 313 SDAGLTHPVQTSIMYVAAAEHYEIVDFEPAYAGOTID-----LRNFAKANGICTDDDYA 367  
QY 519 -----IATDTVANMSSGPEPPFFKNLIASLIPNIADRAPIGKHFVLPAPKPEYR 569  
Db 368 NTDKVMRFHVSQGVVNSVYVPEQ-----LSQIQPPADKTDIDHHR--FHRTNGEMR 418  
QY 570 VGEY-----VEYIVGANPK-----NSAENOTHO-----TELVETKEDSVADMOI 610  
Db 419 INGIGFADVENRYLAKVPRGTVELMELENSGSGSHPIHVLVDFRVVARYGDEGTGVM 478  
QY 611 MY-----NDASWETRFYWHKIGILSNATYWHIPDTAVPGIY-----648  
Db 479 PYEAGLADYVWLGR---HETVL-----VEAHY--APWDGVYFHCNLIHEDQDMAA 527  
QY 649 ----RIRYFGHNKQELLKP 664  
Db 528 FDTYKIQNFGYNETTDEHDP 547  
  
RESULT 4  
US-08-219-262B-9  
Sequence 9, Application US/08219262B  
Patent No. 5788970  
GENERAL INFORMATION:  
APPLICANT: VAKHARIA, VIKRAM  
APPLICANT: SNYDER, DAVID B  
TITLE OF INVENTION: MENDEL-WHERSAT, STEPHANIE A  
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS  
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/219, 262B  
FILING DATE: 29-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2747-047-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1012 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Infectious bursal disease virus

STRAIN: 002-73  
US-08-219-262B-9  
Query Match 3.08; Score 108; DB 1; Length 1012;  
Best Local Similarity 20.98; Pred. No. 0.11;  
Matches 104; Conservative 72; Mismatches 192; Indels 130; Gaps 26;  
  
QY 231 EKNNGVLPGGPFPVAFASS-----NLGDVSPNI-LGPHCVMTGESCDNDK-----275  
Db 153 DKIGNVLVGEIVTLSPITSDIGYVRLGDPIDPAIGLDPKVVATCSDDRPVYTTAAD 212  
QY 276 -----STCPNGSPSCMAAGPGQDMFESTHILIGRIIYOKAKELVYASOEYTGVLAAHQ 330  
Db 213 DYQFSQYQPGGVITLFLSA-NIDAINSLSVGELVQTS-----VQGLVL-----257  
QY 331 VWNKTDVSVOLNATHYKTCRPAIGYSPAGCTIDGCVGLNTGCTGDPFWTLRQOLL 390  
Db 258 --NATIIYLVGFDGT-TVTTTRAVAAAGNGITAGT-DNLMPFNLIPTSEITQPVTSIKLEIV 313  
QY 391 -----GKPSREIVCQPKPILHSGELTI-----PHMQP-DIVDQIVTVGSLAIA 437  
Db 314 TSKSGGAGQDM-----SWLASGNLAVITHGNTPGALRPVTLVAYEVATGVSYYTV 365  
QY 438 AIPGELITMSGRFRREAIKKEFALY--GMKDMTVYI-----AGLSNVY-THYITTYEY 488  
Db 366 AGVSNFELIPNELAKNLVTEYGRDPGAMVYTKLILSERDLGKTYWPFREYTDREY 425  
QY 489 QAQRYEASTYTGHTISAVITQLFRDLAKAIAITVTVANMSSGPEPPFFKNLIASLIPNIA 548  
Db 426 ---FMEVADLNSPLKTAGAG-FKDIIRAIRIAYP-----VSTLFP--464  
QY 549 DRAPIGHFGDVLQPAPEYRVEGEVEYF-----VGANPKSAENOTHOFLYETKEDS 604  
Db 465 PAPILAHIGEV-----DYLGDPAQASSTAASGKARASGRIROLTLADKGYEV 519  
QY 605 VAD-WQIMYNDASWETRFYWHKIGILSNATYWHIPDTAVPGIYRIRYEGHN-----RK 658  
Db 520 VANLFYQPNVY-----DGIL-----ASPVLYR--GAHNDCVLYRE 554  
QY 659 QELLKPAVILAFEGISSP 676  
Db 555 GATLEPVVITTVEDAMTP 572  
  
RESULT 5  
US-09-031-655-9  
Sequence 9, Application US/09031655  
Patent No. 6017759  
GENERAL INFORMATION:  
APPLICANT: VAKHARIA, VIKRAM  
APPLICANT: SNYDER, DAVID B  
TITLE OF INVENTION: MENDEL-WHERSAT, STEPHANIE A  
TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS  
TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031, 655  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/219,262
FILING DATE: 29-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2747-047-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1012 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Infectious bursal disease virus
STRAIN: 002-73
US-09-031-655-9

```

Query Match 3.0%; Score 108; DB 3; Length 1012;

Best Local Similarity 20.9%; Pred No. 0.11; Mismatches 192; Indels 130; Gaps 26;

Matches 104; Conservative 72; Mismatches 192; Indels 130; Gaps 26;

```

QY 231 EKNKGYPGQGPVAVGFASS-----NLGDVSPNI-LGPHCVNTGSCDNDK----- 275
DB 153 DKIGNLVGEGVTLSTPSYDLGVRLGDPPIALGIDPKMVKATCQSSRPRTYTTAAD 212
QY 276 -----STCNGCPSCMASCPCODMFESHTIIGRIITYOKAKELYASASOEVTGPVLAHQ 330
DB 213 DYQFSQYQPGGVITTLFSA-NIDAINSLVSGELVEFQTS-----VOGLVY----- 257
QY 331 WYNNITDVSYOLNATHVKTCKPDLGYSFAAGTIDGVSGLNTGTEGDPFMDTLBDOIL 390
DB 258 --NATIVLVGFGT-TVTRAVAAGNGLTAGT-DNLMPNLVPTSETIOPVTSIKLEY 313
QY 391 -----GRPSEIVECOKPKRPILLHSGELTI-----PHWOP-DIVDQIVTVGSLAIA 437
DB 314 TSKSGQAGDQM-----SWLASGNLAVTHGNYPCALRVTLVAYERAVATGSVTV 365
QY 438 AIPGELTTMGRFRRAIKKEFALY--GKMDTVI-----AGLSNV-THKITTYEEX 488
DB 366 AGVSNFELLPNBELAKNLVTEYGRDPGMNNTKLLSRDLGITVWPTREYTDREY 425
QY 489 QAQRREASTIYCPHTLAVIQLFRDLAKAIAITDVANMSSGPEPEFKNLILASLIPNIA 548
DB 426 -----FMEVADLNSPLKIAAGF-FKDIIRIARIAP-----VSTLFP----- 464
QY 549 DRAPIGKHEGVDLQPAKPEYRVEVEVIEF---VGANPKNSAENOTQOTFLTEKYEES 604
DB 465 PAAPLHAHAGEGV-----DYLGDENQAASGTARAASGRARISROTLAADKGYEV 519
QY 605 VAD-WQIMVNDASWETREFYMHKGIGLSNATYIMHPDPAVPGIYIRIRFGHN-----RK 658
DB 520 VANTLEQVDPNPVY-----DGIL-----ASPGVLR--GAHNLDCVLR 554
QY 659 QELKPAVLAIEGSISSP 676
DB 555 GATLEPPVITTVEDAMTP 572

```

RESULT 6  
US-08-570-311-14

Sequence 14, Application US/08570311

Patent No. 5824791

GENERAL INFORMATION:

APPLICANT: Prologis-Ke-Fox, Ann

APPLICANT: Tumwasorn, Somying

APPLICANT: Lepline, Guyelaine

APPLICANT: Hen, Nalinang

APPLICANT: Lantiz, Marilyn

```

APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSER: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UP15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2628 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-14

```

Query Match 3.0%; Score 106.5; DB 2; Length 2628;

Best Local Similarity 18.8%; Pred No. 0.94; Mismatches 236; Indels 183; Gaps 31;

Matches 116; Conservative 82; Mismatches 236; Indels 183; Gaps 31;

```

QY 101 PAGE--FOYTYTILASEGSNRTQYIVSGIMK-----SIDIAHTNLKPKGFI----- 147
DB 105 PAGLVDPFEYKVPVNDASFSPTNF--VLDGTAADIPAGTYDVIYINHPGIYVGBG 162
QY 148 -NKGNAVAVQINRSPSYLINPOSERARISSNTDKEMLVKLVLDNGELGLISMFALHP 206
DB 163 VSKGNDYVEEAK--TYHFTVORQGP-----GDAAVY----- 193
QY 207 VSMNNSNHFVNSDNGYAAVLEFEQENKKGITLPGQGFVAGFASSNLGDVSPNIGPHCVN 266
DB 194 VTGEGGNEFAPVONLWSV-----SCQTVTLTWQAPASDKRTYVL----- 233
QY 267 TGESCDNDKSTCPNGSPSCMASGPODMFESTHIIGRIITYOKAKELYASASOEVTGPVL 326
DB 234 -NESFDT--QTLPGNG-TMIDADGCGHMLSTIN-----YVTAHTGNG-AM 276
QY 327 AAHQVNMITDVSYOLNATHVTKCKPA-----LGYSFAGTIIDVS---GLNITQCTTE 377
DB 277 FSKSWTASGAKIDISPDNVLVTPKVTVPENCKLSY-WVSSQVPTNHEHYGFLSTGNE 335
QY 378 GDPFMDTLBDOILGKRSSEIVECOKPKPI-LHSGELTTPHWPQDIYDVQIVTVGSLAI 436
DB 336 AANFTIKLEETLGS-----DKPAPNLVKSQGVKLPAVQERTID-----L 377

```





```
REFERENCE/DOCKET NUMBER: B0801/7023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-429-2

Query Match      2.9%; Score 104; DB 4; Length 878;
Best Local Similarity 18.6%; Pred. No. 0.23;
Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

OY 47 ADPOGSMFAVSVELCMISQRLREY-----LKR-----LESYXG 82
DB 41 APPEKRPRLGRVNEEDCTGRQRTAIFLPKVGTDGYITVKRPLRFHPNPTDPFLGLRMD 100
OY 83 SLYRR--DNVILSAIHTHSGPAGFEQYLYLILASEGSENRFTFOYIVSGIMKSIDIAHTNL 140
DB 101 STYKRFSTKVTLLNVTGHHRRP-----PHQASVSGIOAEL-LTFPNS 141
OY 141 KPGKFIKGNVAVNOIRSPSYLLN---POSERARY-----SNTDKEMLYLKV 189
DB 142 SPG-----LRRQRDWVIPISCPENKGFPRKLVQIKSNKKEGVFYSI 188
OY 190 DLNAGEDLGLISMFAL-----HPVSNNNSNHFVNSDNMGYA 224
DB 189 TGCGADTPPVGVFIETRETGLKTEPLDRERATYTLFSAVSSNGN---AVEDPMEL 245
OY 225 AYLFEOEKNK-----GYLPQGPVAVGAFASNLGDVSPNLLGPHCVNTGES 270
DB 246 ITVTDONNKREFTQEVFKGSVMGALPG-----TSVMEV-----TATD 284
OY 271 CDNKSCTPCNGSPSCMASG---PGQDMF---ESTHIGRIITYOKAKELASASQEVYTG 323
DB 285 ADDDVNTY-NAIAVTTILSDPELPDKMFTINRRTGVISVTTGLDRESPTTYTLVVQA 343
OY 324 PVLAHQVNMNTDVSVOLNATHVTKCPALGYSFAAGTIDG-----VSGLNTQ 373
DB 344 ADLOG-EGSLTTATAV---ITVTDNDNRP---FNPTTYGQVPEANAVIITLAKYTD 396
OY 374 GTTGDDPMDTLRDQLGKPSSEIVECOKP-----KPIILHSGELTIPHPWQDI 423
DB 397 ADAPNTPAMEAVY-TILNDGCGFVVTNPNVNDGILKTAKGLDFEAKQVILHVAVTNV 455
OY 424 VDVOIVYVGSIAIAIPBELITMGRFR-----REAIKKEFALYGMKMTVYIAGLSVY 478
DB 456 VPFEVSLTSTAATV-DVLVDNCGPIFVPEKREVSSEDFGV-GOEITSYTAQEPDTF 512
OY 479 THYITTEYEOAQRREASTIYGPH--LSAYIOLFR-----LAKAIAITDVYAN 526
DB 513 MEQKITIYIMWDRNMLE---INPDGAISTRALDRDFEHVKNSTYALILINDNSP 569
OY 527 MSSGPEPPFFKNLIASLIPNIADRAPIGKHGVDYLPQAPKPEYRGEVEVEYIVGANKP-- 584
DB 570 VATG-----TGLTLLILSDVNDNAPI-----PEPRT-----IFECERNRPQ 606
OY 585 -----NSAENQTHQFTLYEKEKEDSVADQIMYNDASWET 619
DB 607 VINIHADLPNPTSPFTALTHGRVPMWTIOYNDPTQES 645

RESULT 9
US-09-798-267-2
Sequence 2, Application US/09798267
Patent No. 6406870
GENERAL INFORMATION:
APPLICANT: Brenner, Michael
APPLICANT: Ceppek, Karyn
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
```

```
TITLE OF INVENTION: Interactions with T Lymphocytes
FILE REFERENCE: L0560/7008ERP
CURRENT APPLICATION NUMBER: US/09/798,267
CURRENT FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 08/237,919
PRIOR FILING DATE: 1994-05-03
PRIOR APPLICATION NUMBER: PCT/US 95/05518
PRIOR FILING DATE: 1995-05-03
PRIOR APPLICATION NUMBER: US 08/732,429
PRIOR FILING DATE: 1996-11-01
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 878
TYPE: PRT
ORGANISM: Homo sapiens
US-09-798-267-2

Query Match      2.9%; Score 104; DB 4; Length 878;
Best Local Similarity 18.6%; Pred. No. 0.23;
Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

OY 47 ADPOGSMFAVSVELCMISQRLREY-----LKR-----LESYXG 82
DB 41 APPEKRPRLGRVNEEDCTGRQRTAIFLPKVGTDGYITVKRPLRFHPNPTDPFLGLRMD 100
OY 83 SLYRR--DNVILSAIHTHSGPAGFEQYLYLILASEGSENRFTFOYIVSGIMKSIDIAHTNL 140
DB 101 STYKRFSTKVTLLNVTGHHRRP-----PHQASVSGIOAEL-LTFPNS 141
OY 141 KPGKFIKGNVAVNOIRSPSYLLN---POSERARY-----SNTDKEMLYLKV 189
DB 142 SPG-----LRRQRDWVIPISCPENKGFPRKLVQIKSNKKEGVFYSI 188
OY 190 DLNAGEDLGLISMFAL-----HPVSNNNSNHFVNSDNMGYA 224
DB 189 TGCGADTPPVGVFIETRETGLKTEPLDRERATYTLFSAVSSNGN---AVEDPMEL 245
OY 225 AYLFEOEKNK-----GYLPQGPVAVGAFASNLGDVSPNLLGPHCVNTGES 270
DB 246 ITVTDONNKREFTQEVFKGSVMGALPG-----TSVMEV-----TATD 284
OY 271 CDNKSCTPCNGSPSCMASG---PGQDMF---ESTHIGRIITYOKAKELASASQEVYTG 323
DB 285 ADDDVNTY-NAIAVTTILSDPELPDKMFTINRRTGVISVTTGLDRESPTTYTLVVQA 343
OY 324 PVLAHQVNMNTDVSVOLNATHVTKCPALGYSFAAGTIDG-----VSGLNTQ 373
DB 344 ADLOG-EGSLTTATAV---ITVTDNDNRP---FNPTTYGQVPEANAVIITLAKYTD 396
OY 374 GTTGDDPMDTLRDQLGKPSSEIVECOKP-----KPIILHSGELTIPHPWQDI 423
DB 397 ADAPNTPAMEAVY-TILNDGCGFVVTNPNVNDGILKTAKGLDFEAKQVILHVAVTNV 455
OY 424 VDVOIVYVGSIAIAIPBELITMGRFR-----REAIKKEFALYGMKMTVYIAGLSVY 478
DB 456 VPFEVSLTSTAATV-DVLVDNCGPIFVPEKREVSSEDFGV-GOEITSYTAQEPDTF 512
OY 479 THYITTEYEOAQRREASTIYGPH--LSAYIOLFR-----LAKAIAITDVYAN 526
DB 513 MEQKITIYIMWDRNMLE---INPDGAISTRALDRDFEHVKNSTYALILINDNSP 569
OY 527 MSSGPEPPFFKNLIASLIPNIADRAPIGKHGVDYLPQAPKPEYRGEVEVEYIVGANKP-- 584
DB 570 VATG-----TGLTLLILSDVNDNAPI-----PEPRT-----IFECERNRPQ 606
OY 585 -----NSAENQTHQFTLYEKEKEDSVADQIMYNDASWET 619
DB 607 VINIHADLPNPTSPFTALTHGRVPMWTIOYNDPTQES 645

RESULT 10
US-09-798-267-3
```



QY 324 PYLAHQVNMVDVSVQLNATHVTCKKALGYSFAGTIDG-----VSGINTQ 373  
DB ADLOG-EGISTATAV---ITVDTNDNPI---FNPITYKGVPENEANVTTLTKVD 396  
QY 374 GTTEGDFWDLRLDOLGKPSSEIYECOR-----KPIILHSGELTIPHPQPI 423  
DB 397 ADAPNTPAWEAVY-TILNDGCGFVTTNPVNDGILKTAKGLDFEAKOOYLHVAVTN 455  
QY 424 VDQVITVGSLSIAIPGELTITMSGRF-----REAIKKEFALYKMDTIVAGLSNY 478  
DB 456 VPEVESLTSTATVY-DVLVDNEGFIYFPEKREVEVSDFG--GGLTSTYAGPDPF 512  
QY 479 THYITTEYEQOQRYAASTIYGPRT--LSATIQLEFRD-----LAKAIATDVAN 526  
DB 513 MEOKITRYIRWDRNMLE---INPDGAISTRALREDDEHVKNSYALAIATDNGSP 569  
QY 527 MSSGPEPPEFKMLIASLIPINADRAPIGHFQDVLQPAKPEYKGVVEVIFGANKP-- 584  
DB 570 VAFG-----TGLLILISDVNDNFI-----PEPRT-----IFCERNPKPO 606  
QY 585 ---NSAENQHOTELTYEKYEDSVADMOIMYNDASWET 619  
DB 607 VINIHADLPMTSPETAELTHGRVPMNTIOYNDPTQES 645

RESULT 12  
US-09-324-867-2  
Sequence 2, Application US/09324867A  
Patent No. 6251632  
GENERAL INFORMATION:  
APPLICANT: Lilly Corp, David  
APPLICANT: Cameron, Cherie  
APPLICANT: Horrocks, L. Suzanne  
APPLICANT: Hough, Christine  
TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use  
FILE REFERENCE: 1669.0010002/7AG/BID  
CURRENT APPLICATION NUMBER: US/09/324,867A  
CURRENT FILING DATE: 1999-06-03  
EARLIER APPLICATION NUMBER: 09/035,141  
EARLIER FILING DATE: 1998-03-05  
EARLIER APPLICATION NUMBER: 60/039,953  
EARLIER FILING DATE: 1997-03-06  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 2343  
TYPE: PRT  
ORGANISM: Canis familiaris  
US-09-324-867-2

Query Match 2.9%; Score 104; DB 4; Length 2343;  
Best Local Similarity 17.2%; Pred. No. 1.4;  
Matches 144; Conservative 102; Mismatches 295; Indels 296; Gaps 37;  
QY 87 RDNVILSAHTHSAGPGFYQTYLIASE---GFSNRTFYIYSGIMKSIDIANT---- 138  
DB 1214 QENVALLPOAHTMIGTKNFKL-NLFLSTKONVAGLEEQPYTPILOTRLSNDSPEGTH 1272  
QY 139 -----NLKKGKIFINKGNVANOINSPPSYLLNPOSERRARYSNTDKEL--- 184  
DB 1273 MANFSKIREANLE-----GLGNOTNOMVERPST-----TRMSSNSAQHITOR 1317  
QY 185 -----VLKLVLDNGEDLGLISWPAI 204  
DB 1318 GKRLSKQRLSGGEIKFERKVIANDSTQWKMNMNYLAGCTLTQLEYNKEKRAITQSL 1377  
QY 205 HPSVSMNN-----SNHVSNDNMGYAA----- 225  
DB 1378 SDCSKRNHVTIOMNSALPVAKEASPSVRYHTDLTKIPSOH--NSSHLPASACNYFRER 1435  
QY 226 -----YLFEQEKNK-----GYLPGGGPFVAGPASSNLGVSNN----- 258

DB 1436 TSGVQESHFLQEAQRNNLSLAEVTLTGTEGQKF-----SSLGKSATNPMYKLENT 1489  
QY 259 -ILGPHCVMTGES-----CDNDKSTPCNGSPSCMASGPGODMFESTHIGRIYOK 309  
DB 1490 VILQPOLSETSDKVELLSQVNHQDEDSF-----PRTSNDSGH-----LDLMGKTFLOK 1539  
QY 310 AK-----ELYASASQEVTPYLAHQVNMVDVSV-----OL 341  
DB 1540 TCGPVAMKNTNSPGKPFKWMATESSEKIPSKILGLAVLADNHNHYDQIPSEEWKQKSOQ 1599  
QY 342 NATHYVTKQKPLALGYSFAAGTIDGVSGLNITGTEGDFWDLRLQOLGKPSSEIVEGO 401  
DB 1600 NTAFKKRDITLPLG---PCENNDSTAINEGDKPREAMW-----AKGEGRLCS 1648  
QY 402 KKPILL-HSGELTIPHPQPIVDQVITVGSLSIAA---IPGELTITMSGRFPEAIK 456  
DB 1649 QNRPVSKNHQRETTV-TIQPEDKREYDQTSIEKRRDPIDYGYENQGLRSQKTR 1707  
QY 457 KEFAL-----YGMKMTVTI---AGLSNVYTHYITTEYEQOQRYEASTIYGPHTLS 506  
DB 1708 HYFIAVERLWPMYGMKSRPHILRNRAOSGDVOQFKKVVQEFPTDGSF--TOPLYRGE-LN 1764  
QY 507 AYIQLEFRDLAKAIATD---TVANNSSGPEPPEFKMLIASLIPINADRAPICK----- 555  
DB 1765 EHLGLQPYTRAEVEDNINIVYTERKNOASRPY-SFYSSLISYDEDEGOGAPRRKFNPNFT 1823  
QY 556 --HFGDV--LOPAKPEY-----RVGEVVEYIFGANKPSAENO--T 591  
DB 1824 KIYFMKVQHNMNAPTKOEFDCKAMAYFSDVLDKDNHSGILGLTLCRSWTLNPAHGRQVT 1883  
QY 592 HQTFLTYEKYEDSVADMOIMY-----DASWETRFYWH-----KCI 627  
DB 1884 VOEFALVFTIFDSTKSWYFTENLERNCRAPCVOKEDPLKCNFRHAINGVYKDTLPGL 1943  
QY 628 LGLSNATYWHIPDT-AVPGIRIRYFGH---NRKOELLKPAVILAPEGISSPFEV 680  
DB 1944 VAAQDKVAMVYLLSMGSENENHSHFSGHVTFYRKKEEYKAMVNYLIPGV---FETV 1997

RESULT 13  
US-08-989-299-12  
Sequence 12, Application US/08989299  
Patent No. 6194556  
GENERAL INFORMATION:  
APPLICANT: Acton, Susan L.  
APPLICANT: Robinson, Keith E.  
TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG.  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/989,299  
FILING DATE: 11-DEC-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold E., Beth  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: MIA-025.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 907 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-989-299-12

Query Match 2.8%; Score 100.5; DB 4; Length 907;  
Best Local Similarity 18.7%; Pred. No. 0.58; Indels 219; Gaps 29;  
Matches 116; Conservative 72; Mismatches 213;

134 DIAHT--NLKPGK-----FINKGVANVOINRSPSYLLNPOSRARYSN--178  
107 DVAFSTASTDDELTEVIEIOLVPLNTGSIASNKTKGP--VFAPVAQALVNSSYWM 164  
179 -----TKEMLVKLVLDNGE-----DLGLISWFAIHVVS-----MNN 211  
165 KTDMLQAPGISIKDEKLRSLWLAGYEAEIKVLRVALSGWRYFNDASPSLKALDEAENV 224  
212 SNHFNVDNMGYAYLFEQ-----KNGYLPGCGPFYAGASMSLGDVSPNIIIGPH 263  
225 LTFEVRKSTSM--QAKQFDMASVTDEKYNRQLGVVSFEQ--MSALAPSRFADYSQ--AQA 277  
264 CVNTGESCDNDKSTCPNGSPSCMASGPGQD--MESTHIIIRIITY-----QAKEL 313  
278 ALNND-----SKDSTICDMDVPPCALQKIDMSIFRNEKDSRLQHLWVSYYTALAKSPS 334  
314 Y-----ASASQEVTVGLAHLAHOVNMNVDVSVQLNAT-----H 345  
335 YNNITITISNEGAKINGFANGAMWRSADFDMSSKVKHKAFFDLNKOIIRKYSTIOPFYQLH 394  
346 TVKCKPAGLGSFAG--TIOGVSLNITGTEGDPMDLRLDQLAKPSSEIYECQPK 404  
395 AYMRRLAGTISNPGVSKDGPIDPHLGLSDGD--WSAHYEQT--KPEEE--ESETPE 448  
405 PILHSGELTIPHPWOPDIVVOIVTVGSLAIAIPGELTMSGRFREFA-----454  
449 AML-----SAFTQNTYTKKMFVTAY-----RYEFSAFPHLPKSYW 485  
455 IKKEFALYKMDTFVVIAG-----LSNVYHYITTYEYEQAR 492  
486 TSSIFARWYSKDMICHPAALDMRAPNDFRYKACAQLGEPPEDQASHLVQTYOY---541  
493 YEASTIYPTLSAYIQLPRDLAKATATDVAN---MSSGPEPFEPKNIASLIPNIA 548  
542 -----LKQDQSL-----LFEQASPVITDAIANAFALSTNPHYLVSQKLVSEHLDIK 590  
549 DRAPIGHFGDVLDP-----AKPEYRGEVEVEIVFGANPKISAENQTHQTFLEVEK 600  
591 DSVIINKLYKESLESFRLKPTTIAADNMRY-----ELFDGTVPKPK-----631  
601 YEDSVADWQIMYNDASWETR 620  
632 -----LNDRWWEIR 640

RESULT 14  
US-09-026-001A-10  
Sequence 10, Application US/09026001A  
Patent No. 6413760  
GENERAL INFORMATION:  
APPLICANT: Boonhoo, Amechand  
APPLICANT: Seehra, Jasbir  
APPLICANT: Shaw, Gray  
APPLICANT: Sako, Dianne  
TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM  
TITLE OF INVENTION: THERAPEUTIC USES THEREOF  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/09/026, 001A  
APPLICATION NUMBER: US/09/026, 001A  
FILING DATE: 18-FEB-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G15293B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 613 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-026-001A-10

Query Match 2.8%; Score 100; DB 4; Length 613;  
Best Local Similarity 23.7%; Pred. No. 0.32;  
Matches 54; Conservative 33; Mismatches 103; Indels 38; Gaps 10;

85 YRRDNVIL-----SAIHHSRPGAFQYTYLIIASSEGSRNTFYISGIMKSIDI-----135  
221 YKRQKPVYKRRYEMINIMNMYNRLNHLALIGLEIWSNNEINVOVDVATDLRGEW 280  
136 AHTNLKPGKIPINKGVANVOINRSPSY-----LNPQSRFA--RYSSWTDEKMLVLK 187  
281 REKLLPRKRDNDMLGIDFGKTPVGLAYIGSICNKSVAAYVDYSST--SWAIT 338  
188 LVLDNGEDLGLISWFAIH---PVSMMNSNHFVNSDNMGYAYLFEQ---EKNGYLPQG 241  
339 MAHEHGNMG-----IHHDGPTSCGSKNCVMSTRTPEPAYQFSSCSVREHQLLDR 392  
242 PFVAGFASNSLGDVSPNIIIGPHCVNTGESCDNDKSTCPNGSPSCMAS 289  
393 PQCILNKLPLSDIVSPICGNNFVEYGECD-----C--GSPADCQA 433

RESULT 15  
US-09-026-001A-18  
Sequence 18, Application US/09026001A  
Patent No. 6413760  
GENERAL INFORMATION:  
APPLICANT: Boonhoo, Amechand  
APPLICANT: Seehra, Jasbir  
APPLICANT: Shaw, Gray  
APPLICANT: Sako, Dianne  
TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM  
TITLE OF INVENTION: THERAPEUTIC USES THEREOF  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/026,001A  
 FILING DATE: 18-FEB-1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BROWN, SCOTT A.  
 REGISTRATION NUMBER: 32,724  
 REFERENCE/DOCKET NUMBER: G152938  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 498-8224  
 TELEFAX: (617) 876-5851  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 621 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-026-001A-18

Query Match 2.8%; Score 100; DB 4; Length 621;  
 Best Local Similarity 23.7%; Pred. No. 0.32;  
 Matches 54; Conservative 33; Mismatches 103; Indels 38; Gaps 10;

OY	85 YRDNVIL-----SAITHSGPAGFFQYTLVLIASEGFSNRTFOYIVSGIMKSIDT-----135
DB	229 YRDRPVIRRYVEMINTMNNVYRLNLFIALIGLEIWSNRNEINVOQDQATLDFGEW 288
OY	136 AHTNLPKPKIFINKGVANVQINRSPSY-----ILNPOSERA---RYSSNTDKEMLVLK 187
DB	289 REKKLLPKRRNDNAQLTGIDKRGTPVGLAYIGSICNPKSSVAVVODYSRT--SWVAIT 346
OY	188 LVDLNGEDGLISWFAIH---PVSNNNSNHFVNSDNMGYAAVLFEO---EKNKGYLPCOG 241
DB	347 MAHEMGHNG-----IHHDGPSCCTCGSNKCVWSTRTEPAYQFSSCSVREHQEYLLRDR 400
OY	242 PVVAGFASNLGADVSPNITLGPHCVNTGESCNDNDKSTCPNGGPGSMCMAS 289
DB	401 POCILNKPPLSTDIVSPPIGNNFVEVGECD---C--GSPADCOSA 441

Search completed: July 3, 2003, 12:53:08  
 Job time : 30 secs

5'143841  
 4'997,760

GenCore version 5.1.6.  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 3, 2003, 12:39:23 ; Search time 73 Seconds

(without alignments)  
1244.889 Million cell updates/sec

Title: US-09-937-521-14

Perfect score: 3582  
Sequence: 1 FSGYIGVGRADCTGVSDI.....KPAVLAEGISPFVVT 682

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1931.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3582	100.0	682	21	AAB33292	Mouse mature neutr
2	3582	100.0	756	21	AAB33291	Mouse neutral/alka
3	3312	92.5	745	22	AAB97029	Rat protein involy
4	3054	85.3	761	23	AAE13544	Human ceramidase p
5	1509.5	42.1	704	22	ABE58152	Drosophila melano
6	1090.5	30.4	646	21	AAV57307	P. aeruginosa matu
7	1090.5	30.4	670	21	AAV57314	P. aeruginosa cera
8	879	24.5	280	20	AA138464	Human secreted pro
9	167	4.7	35	21	AAB33290	Mouse neutral/alka
10	148	4.1	80	21	AA603568	Human secreted pro

11	146.5	4.1	1565	18	AAW18305	Photorhabdus lumin
12	131.5	3.7	1565	19	AAW56568	Toxin TcoB, encode
13	127	3.5	27	20	AAV38468	Human secreted pro
14	125	3.5	26	20	AAV38466	Human secreted pro
15	118.5	3.3	2013	23	ABBA4751	Listeria monocytog
16	116.5	3.3	766	22	AAU34979	Enterococcus faeca
17	116	3.2	1339	21	AAAB42887	Human ORFX ORF2651
18	116	3.2	1355	22	AAAB1859	Human polypeptide
19	113	3.2	1981	19	AAWA2634	Protein sequence t
20	112	3.1	812	22	AAAO0932	Human polypeptide
21	111.5	3.1	627	21	AAV9538	Bipolaris spicifite
22	111.5	3.1	627	21	AAV96762	Bipolaris spicifite
23	111.5	3.1	627	23	AAO18211	B spicifite phenol
24	111.5	3.1	952	22	AAW78715	Human protein SEQ
25	109.5	3.1	582	20	AAV23920	Amino acid sequenc
26	109.5	3.1	699	20	AAV08471	F. balustinum CP70
27	109	3.0	629	22	AAV34412	Novel human secret
28	109	3.0	752	22	AAW39146	Human polypeptide
29	109	3.0	1978	20	AAV27230	Amino acid sequenc
30	107.5	3.0	937	22	AAW78714	Human protein SEQ
31	106.5	3.0	2628	17	AAW6030	P. gingivalis haem
32	106.5	3.0	2628	19	AAW69488	Haemagglutinin pro
33	106	3.0	1021	7	AAV61316	Infectious bursal
34	104	2.9	878	15	AAW55060	Sequence of human
35	104	2.9	878	16	AAW85487	Human E-cadherin p
36	104	2.9	878	23	ABW81475	Human E-cadherin p
37	104	2.9	1352	22	ABG08357	Novel human diagno
38	104	2.9	2343	20	AAW80989	Canine factor VIII
39	104	2.9	2343	21	AAV57846	Canine Factor VIII
40	103.5	2.9	914	22	ABW63634	Drosophila melano
41	103.5	2.9	1026	22	ABW61775	Drosophila melano
42	103	2.9	1413	18	AAW20725	H. pylori secreted
43	102.5	2.9	1026	21	AAV33025	Staufen protein of
44	101.5	2.8	865	22	AAV95798	Human protein sequ
45	101.5	2.8	865	22	AAW67610	Amino acid sequenc

#### ALIGNMENTS

RESULT 1	
AAB33292	standard; Protein: 682 AA.
XX	
AC	AAB33292;
XX	
DT	02-FEB-2001 (first entry)
XX	
DE	Mouse mature neutral/alkaline ceramidase, SEQ ID NO:14.
XX	
KW	Neutral/alkaline ceramidase; mouse; murine; recombinant production;
KW	cellular ceramide content control; antibody; membrane synthesis;
KW	lipid engineering; ceramide metabolism; drug development;
KW	mature ceramidase;
OS	Mus sp.
XX	
PN	W0200058448-A1.
XX	
PD	05-OCT-2000.
XX	
PF	24-MAR-2000; 2000WO-JP01802.
XX	
PR	26-MAR-1999; 99JP-0084743.
PA	(TAKI ) TAKARA SHUZO CO LTD.
XX	
PI	Ito M;
XX	
DR	WPI: 2000-619079/59.
DR	N-PSDB: AAA97634.
XX	
PT	Mammalian neutral/alkaline ceramidase applicable as lipid engineering

PT reagent for studying structure and functions of ceramide as well as  
PT developing drugs for diseases associated with ceramide metabolism  
XX  
XX  
PS Claim 1: Page 62-65; 76pp; Japanese.

CC The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents the mature mouse  
CC neutral/alkaline ceramidase.

XX  
XX  
SQ Sequence 682 AA:

Query Match 100.0%; Score 3582; DB 21; Length 682;  
Best Local Similarity 100.0%; Pred. No. 6.8e-313;  
Matches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLTLRFSAFIADPDGSRMAFVS 60  
DB 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLTLRFSAFIADPDGSRMAFVS 60

QY 61 ELCMISQRLRELVKRLKESKYGSLYRRDVIISAITHSGPAGFFQYTYIILASEGFSNR 120  
DB 61 ELCMISQRLRELVKRLKESKYGSLYRRDVIISAITHSGPAGFFQYTYIILASEGFSNR 120

QY 121 TFQYIYSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSSND 180  
DB 121 TFQYIYSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSSND 180

QY 121 TFQYIYSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSSND 180  
DB 121 TFQYIYSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSSND 180

QY 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAALFDEKKKGYLPQ 240  
DB 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAALFDEKKKGYLPQ 240

QY 241 GPFVAGFASSNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300  
DB 241 GPFVAGFASSNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300

QY 301 IIGRIYOKAKELIYASASQEVTPVLAHQWVMNTDVSQVLAHTHTVCKPALGYSFAA 360  
DB 301 IIGRIYOKAKELIYASASQEVTPVLAHQWVMNTDVSQVLAHTHTVCKPALGYSFAA 360

QY 361 GTIDGVSGLNTGTEGDPFMDTLRDOLGKPSSEIVECOKKPIILLHSGELTIHPMQ 420  
DB 361 GTIDGVSGLNTGTEGDPFMDTLRDOLGKPSSEIVECOKKPIILLHSGELTIHPMQ 420

QY 421 PDIVDQIYTVGSLAIAIPGELTTWSGRRFRAIKKEFALYKMDTVIAGLSNVYTH 480  
DB 421 PDIVDQIYTVGSLAIAIPGELTTWSGRRFRAIKKEFALYKMDTVIAGLSNVYTH 480

QY 481 YTTTVEYQAOQRYEASTIYCPHTLSAYIQLFRDLAKAIATDTVANMSSGPEPPFKNLI 540  
DB 481 YTTTVEYQAOQRYEASTIYCPHTLSAYIQLFRDLAKAIATDTVANMSSGPEPPFKNLI 540

QY 541 ASLIPDIADRAPIGKFGDVLOPAKPEYVGEVEVEIFGANKNSAENOTHTETLVEK 600  
DB 541 ASLIPDIADRAPIGKFGDVLOPAKPEYVGEVEVEIFGANKNSAENOTHTETLVEK 600

QY 601 YEDSVADQIMYNDAWETRFYWHKGIILSNATIYWHIPDTAIPGIYIRYEGHNRKOE 660  
DB 601 YEDSVADQIMYNDAWETRFYWHKGIILSNATIYWHIPDTAIPGIYIRYEGHNRKOE 660

QY 661 LKPAVILAFEGISSPEFEVYTT 682  
DB 661 LKPAVILAFEGISSPEFEVYTT 682

RESULT 2

AAB23291  
ID. AAB23291 standard; Protein; 756 AA.

XX  
XX  
AC AAB23291;

DT 02-FEB-2001 (first entry)

XX  
XX  
DE Mouse neutral/alkaline ceramidase, SEQ ID NO:13.

XX  
XX  
KW Neutral/alkaline ceramidase; mouse; murine; recombinant production;

XX  
XX  
KW cellular ceramide content control; antibody; membrane synthesis;

XX  
XX  
KW lipid engineering; ceramide metabolism; drug development.

XX  
XX  
OS Mus sp.

XX  
XX  
PN WO200058448-A1.

XX  
XX  
PD 05-OCT-2000.

XX  
XX  
PF 24-MAR-2000; 2000MO-JP01802.

XX  
XX  
PR 26-MAR-1999; 99JP-0084743.

XX  
XX  
PT (TAKI ) TAKARA SHUZO CO LTD.

XX  
XX  
PI Ito M.

XX  
XX  
DR WPI: 2000-619079/59.

XX  
XX  
DR N-PSDB: AAA97633.

PT Mammalian neutral/alkaline ceramidase applicable as lipid engineering  
PT reagent for studying structure and functions of ceramide as well as  
PT developing drugs for diseases associated with ceramide metabolism  
XX  
XX  
PS Example 4; Page 58-62; 76pp; Japanese.

CC The invention relates to a mouse neutral/alkaline ceramidase and to  
CC nucleic acids encoding it. The invention also relates to expression  
CC vectors and host cells comprising DNA encoding mouse neutral/alkaline  
CC ceramidase, the recombinant production of the ceramidase, an antibody  
CC against the ceramidase, and a method for controlling the ceramide  
CC content in cells and/or tissues by introducing the gene or its antisense  
CC nucleic acid into the cells and/or tissues. The ceramidase can be used  
CC as a lipid engineering reagent for studying structure and functions of  
CC ceramide as well as developing drugs for diseases associated with  
CC ceramide metabolism. The present sequence represents the mouse  
CC neutral/alkaline ceramidase.

XX  
XX  
SQ Sequence 756 AA:

Query Match 100.0%; Score 3582; DB 21; Length 756;  
Best Local Similarity 100.0%; Pred. No. 8.1e-313;  
Matches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLTLRFSAFIADPDGSRMAFVS 60  
DB 1 FSGYITGVGRADCTGVSDINLMGYKNGONAGLTLRFSAFIADPDGSRMAFVS 60

QY 75 FSGYITGVGRADCTGVSDINLMGYKNGONAGLTLRFSAFIADPDGSRMAFVS 134  
DB 75 FSGYITGVGRADCTGVSDINLMGYKNGONAGLTLRFSAFIADPDGSRMAFVS 134

QY 61 ELCMISQRLRELVKRLKESKYGSLYRRDVIISAITHSGPAGFFQYTYIILASEGFSNR 120  
DB 61 ELCMISQRLRELVKRLKESKYGSLYRRDVIISAITHSGPAGFFQYTYIILASEGFSNR 120

QY 121 TFQYIYSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSSND 180  
DB 121 TFQYIYSGIMKSIDIAHTNLKPKKIFINKGNVANOINSPSSYLLNPOSERARYSSND 180

QY 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAALFDEKKKGYLPQ 240  
DB 181 KEMLVKLVLDNGEDGLISWFAIHVSNMNSHFNPSNDMGYAAALFDEKKKGYLPQ 240

QY 241 GPFVAGFASSNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300  
DB 241 GPFVAGFASSNLGDVSPNLTGPHCVNTGSCDNKSTCPNGGSPKMGASGPGDMEESTH 300



QY 301 IIGRIYQAKELASASQEVYTGPIAAHQVNMNTDVSVOLNATHVTYCKPALGYSFAA 360  
 Db 375 IIGRIYQAKELASASQEVYTGPIAAHQVNMNTDVSVOLNATHVTYCKPALGYSFAA 434  
 QY 361 GTIDVSGSLNTGCTTGBDPFMDTLRDQLCKPSEIEVECKPKPILHSGELTIPHPMQ 420  
 Db 435 GTIDVSGSLNTGCTTGBDPFMDTLRDQLCKPSEIEVECKPKPILHSGELTIPHPMQ 494  
 QY 421 PDIVDQIVTVGSLAIAIPBELTMSGRFRREAIKKEFALYGMKDMTVVAGLSNYTH 480  
 Db 495 PDIVDQIVTVGSLAIAIPBELTMSGRFRREAIKKEFALYGMKDMTVVAGLSNYTH 554  
 QY 481 YITTYEYOQRYEAASITYPHTLSAVIQLEFRLAKAIAIDTVANNSSGPEPPFNLI 540  
 Db 555 YITTYEYOQRYEAASITYPHTLSAVIQLEFRLAKAIAIDTVANNSSGPEPPFNLI 614  
 QY 541 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEIFVGANPKNSAENQTHQFLYVEK 600  
 Db 615 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEIFVGANPKNSAENQTHQFLYVEK 674  
 QY 601 YEDSVADQIMYNDAWETREYMHKGIIGLSNATYHIDPTAIPGIRIRYFGHNKOE 660  
 Db 675 YEDSVADQIMYNDAWETREYMHKGIIGLSNATYHIDPTAIPGIRIRYFGHNKOE 734  
 QY 661 LKPAVILAFEGISSPEFVVT 682  
 Db 735 LKPAVILAFEGISSPEFVVT 756

RESULT 3  
 AAB97029 standard; protein: 745 AA.  
 XX AAB97029;

XX 20-JUL-2001 (first entry)

DE Rat protein involved in hydrolysis of the acid amide bond in ceramide.

KW Rat; acid amide bond hydrolysis; ceramide; cytostatic; cancer.

OS Rattus norvegicus.

XX P11332P00105/890.FA

XX 06-MAR-2001

XX 23-AUG-1999 99JP-0235218.

XX 23-AUG-1999 99JP-0235218.

XX (MITU) MITSUBISHI CHEM CORP.

XX SMP1: 2001-304133/32.

XX N-PSDB: AAF99977.

PT Novel protein, used to treat abnormal cell growth e.g. in cancers, is capable of hydrolyzing the acid amide bond between sphingosine and the fatty acid in ceramide.

PS Claim 1; Page 12-13; 14pp; Japanese.

XX The present sequence is provided in a specification relating to a protein having a 761 residue amino acid sequence or its variant comprising a replacement, deletion, insertion, addition or reversion of at least one amino acid, but retaining the activity of hydrolyzing the acid amide bond between sphingosine and the fatty acid in a ceramide.

CC The protein can be used to treat diseases caused by abnormal cell growth such as cancers.

CC growth such as cancers.

XX Sequence 745 AA;

Query Match 92.5%; Score 3112; DB 22; Length 745;  
 Best Local Similarity 91.8%; Pred. No. 1.66-288;  
 Matches 626; Conservative 25; Mismatches 15; Indels 16; Gaps 1;

QY 1 FSGYIGYGRADCTGQVSDINLMGKGNQARGLIRLEFRATILADPDGSSNMAFVS 60  
 Db 80 FSGYIGYGRADCTGQVSDINLMGKGNQARGLIRLEFRATILADPDGSSNMAFVS 139  
 QY 61 ELCMISORLREVLKRESKYGSILYRDNVTLSTHSSGAGFEQYTLTLASEGSNR 120  
 Db 140 ELCMISORLREVLKRESKYGSILYRDNVTLSTHSSGAGFEQYTLTLASEGSNR 199  
 QY 121 TFOYIVSGIMKSIDIAHTNLKPKIFIKGNVAVQINRSSTYLLNPOSERARVSSNTD 180-  
 Db 200 TFOYIVSGIVASIDIAHTNLKPKGVILINKGNVAVQINRSSTYLLNPOSERARVSSNTD 259  
 QY 181 KEMVLKLVLDNGEDLGLISWFAIHPVSMNSNHFPVSDNNGVAAVLEQEKNGYLPQ 240  
 Db 260 KEMVLKLVLDNGEDLGLISWFAIHPVSMNSNHFPVSDNNGVAAVLEQEKNGYLPQ 319  
 QY 241 GPVAGFASMLGQVSPNIIIGPHCVNTEGSCDNDKSTCPNGSPGCMASGPGDMFESTH 300  
 Db 320 GPVAGFASMLGQVSPNIIIGPHCVNTEGSCDNDKSTCPNGSPGCMASGPGDMFESTH 379  
 QY 301 IIGRIYQAKELASASQEVYTGPIAAHQVNMNTDVSVOLNATHVTYCKPALGYSFAA 360  
 Db 380 IIGRIYQAKELASASQEVYTGPIAAHQVNMNTDVSVOLNATHVTYCKPALGYSFAA 439  
 QY 361 GTIDVSGSLNTGCTTGBDPFMDTLRDQLCKPSEIEVECKPKPILHSGELTIPHPMQ 420  
 Db 440 GTIDVSGSLNTGCTTGBDPFMDTLRDQLCKPSEIEVECKPKPILHSGELTIPHPMQ 499  
 QY 421 PDIVDQIVTVGSLAIAIPBELTMSGRFRREAIKKEFALYGMKDMTVVAGLSNYTH 480  
 Db 500 PDIVDQIVTVGSLAIAIPBELTMSGRFRREAIKKEFALYGMKDMTVVAGLSNYTH 559  
 QY 481 YITTYEYOQRYEAASITYPHTLSAVIQLEFRLAKAIAIDTVANNSSGPEPPFNLI 540  
 Db 560 YITTYEYOQRYEAASITYPHTLSAVIQLEFRLAKAIAIDTVANNSSGPEPPFNLI 619  
 QY 541 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEIFVGANPKNSAENQTHQFLYVEK 600  
 Db 620 ASLIPNADRAPICKHFGDVLQPAKPEYRGEVEVEIFVGANPKNSAENQTHQFLYVEK 663  
 QY 601 YEDSVADQIMYNDAWETREYMHKGIIGLSNATYHIDPTAIPGIRIRYFGHNKOE 660  
 Db 664 YEDSVADQIMYNDAWETREYMHKGIIGLSNATYHIDPTAIPGIRIRYFGHNKOE 723  
 QY 661 LKPAVILAFEGISSPEFVVT 682  
 Db 724 LKPAVILAFEGISSPEFVVT 745

RESULT 4  
 AAE13544 standard; protein: 761 AA.  
 XX AAE13544;

XX 26-FEB-2002 (first entry)

XX Human ceramidase protein.

XX Human; ceramidase; nephrotropic; antipsoriatic; antitumor therapy;

KW gene therapy; proliferative disorder; cancer; cardiovascular disease;

KW inflammation; neurodegenerative disorder; cytostatic; immunosuppressive;

KW nootropic; signal transduction; breast cancer; autoimmune disorder;

KW Alzheimer's disease; growth deficiency; lesion; lupus nephritis;

XX glomerular disease.

XX Homo sapiens.

XX Key Location/Qualifiers

FT	Peptide	1..19
FT	/label- Signal_peptide	20..761
FT	Protein	17..22
FT	Modified-site	/label- Myristoylation_site
FT	Region	37..66
FT	Domain	/note= "Low compositional complexity signal"
FT	Region	176..196
FT	Domain	/label- Transmembrane_domain
FT	Domain	311..327
FT	Domain	/note= "Determined by Edman sequencing of purified rat brain enzyme"
FT	Domain	313..333
FT	Domain	/label- Transmembrane_domain
FT	Domain	431..451
FT	Domain	/label- Transmembrane_domain
FT	Domain	506..525
FT	Domain	/label- Transmembrane_domain
FT	Domain	543..563
FT	Region	/label- Transmembrane_domain
FT	Region	635..648
FT	Region	/note= "Determined by Edman sequencing of purified rat brain enzyme"
FT	Region	737..753
FT	Region	/note= "Determined by Edman sequencing of purified rat brain enzyme"
XX	MO200155410-AZ.	
PN	02-AUG-2001.	
XX	29-JAN-2001; 2001WO-US02866.	
XX	28-JAN-2000; 2000US-178975P.	
XX	(MUSC-) MUSC FOUND RES DEV.	
PA	Hannun YA, El Bawab S;	
PI	WPI: 2002-025687/03.	
DR	N-PSDB: AAD22668.	
XX	Human mitochondrial ceramidase protein and gene, modulation of which is useful for preventing and treating proliferative disorders e.g. cancer, cardiovascular disease, inflammation and neurodegenerative disorders -	
XX	Claim 28; Fig 3; 113pp; English.	
XX	The present sequence is human mitochondrial ceramidase protein.	
XX	Ceramidase protein and gene are useful for treating a disease or disorder involving cell over proliferation or sphingolipid signal transduction especially breast cancer, cardiovascular disorder or inflammation. Ceramidase protein and gene are useful for treating disorders involving deficient cell proliferation or growth e.g. neurodegenerative disorders (Alzheimer's disease), growth deficiencies and lesions. Ceramidase protein is also useful for diagnosis of hyperproliferative diseases. Ceramidase gene can be used as an immunogen to generate antibodies which are useful for diagnosis, prevention and treatment of hyperproliferative diseases and for detecting ceramidase gene product in a biological sample. The hyperproliferative disorders include cancers and autoimmune disorders such as lupus nephritis, glomerular disease. Ceramidase gene is also useful in antisense therapy and gene therapy.	
XX	Sequence 761 AA:	
SO	Query Match 85.3%; Score 3054; DB 23; Length 761;	
	Best Local Similarity 82.5%; Pred. No. 2.7e-265;	
	Matches 562; Conservative 63; Mismatches 56; Indels 0; Gaps 0	
0Y	1 FSGYIGGRADCTGQVSDINLMGCGKGNQANRGLTLRFSAPFLADBDGSGNRATFSV 60	

Dd	80	FSGHHIGVGAADCTQGVADINILMGCKSGSSQNQGLITFLTLYSRAFIIMAPDSSNNTFVSI	139
Oy	61	ELCHISORLRLVLEVKLESRYGSLTRRDVNLISAITHSGPAGFQOTLYTLASEGFSNR	120
Dd	140	DIGVWSORLRLLEVLANLRLOSKYSLYRRDVLNLSGTHHSGAGAGFYQYVFWFVIASEGFSNQ	199
Oy	121	TFQYISGIMKSLIDIAHTNLKPKGKLFINKGNVANYQIRSPSVLNPQSRARYSSNTD	180
Dd	200	TFQHMVGTGLKSLIDIAHTNMKPKGLTFINKGNVDVQVIRSPSYSTQDNPOSERARYSSNTD	255
Oy	181	KEMVLVLTLDNGEDGLISWFAIRHPVSNNNSNHFVNSDNGVYAAAYFEQEKNGYLRPGQ	240
Dd	260	KEMVLVLMVDNLNDGDLISWFAIRHPVSNNSNHLVNSDNGVYASYLLEQKNGGYLRPGQ	319
Oy	241	GPEYVAGFASNLCDVSPNLRGHCYNTGSGCDNDKSTCPNGGSPSCMAAGFGQDMFESTH	300
Dd	320	GPEFAAFASNLCDVSPNLRGRCITNTGSGCDNASTCPIGGSPMCIKAGGQDMFESTQ	379
Oy	301	LIGRIIYQKKELIYASASEVTPGVLAHQVYNNMTDVSVOLNATHYVTCRPAIGYSFAA	360
Dd	380	LIGRAMQQRKELIYASASEVTGPLASAHQVMDMTDVVNLNSTHASTKCPALGYSFAA	439
Oy	361	GTDIGVSGNLITQGTGEGDFPMDTLRDQLGKPSSEIVEECOKPRILHSGELTIPPMQ	420
Dd	440	GTDIGVQGLNFTQGTGEGDFPMDTIRDOILGKPSSEIECKKPRILHTELSELKPRPMH	499
Oy	421	PDIVDVQIVTVYVSLAIAAIPGELITTMGSRGRFREALIKFEPALYGMKDMTVIAGLSNYTH	480
Dd	500	PDIVDVQIITLGLSLITAIIPGETTMMGSRRLREAVQAFASGQMNMVTVISGLCNYTH	559
Oy	481	YITTYEYQARKEASTIYGPHLTSATYQLFRDLAKAIAIDTYAANSSGGEPPFFNLI	540
Dd	560	YITTYEYQARKEASTIYGPHLTSATYQLFRDLAKAIAIDTYAANSSRGEPPFFNLI	619
Oy	541	ASLIPNIAADPARGIKHFEGDLPQAKREYRVEGEVVEIVFGANPKNSAENOTHOTFLVEK	600
Dd	620	VPLIPSLVDRAPRGKTFGVDLPQAKREYRVEGEVAEIVFGANPKNSVONOTHOTFLVEK	679
Oy	601	YEDSVADWDQIMYNDASMETRFYHNKGIQLSNATITMHIPTAAYPGYLRIRYEGHNKOE	660
Dd	680	YEATSTSMQYVCDASMETRFYHNKGIQLSNATVEMHIPTAQQGIRIRIRYEGHNKOD	739
Oy	661	LKRPAYILAEFGISSPFEVYT 681	
Dd	740	ILKRPAYILSFEGSTSPAFEVYT 760	
RESULT 5			
ABB58152	ID ABB58152 standard; Protein: 704 AA.		
XX	ABB58152;		
XX	26-MAR-2002 (first entry)		
XX	Drosophila melanogaster polypeptide SEQ ID NO 1248.		
XX	Drosophila; developmental biology; cell signalling; insecticide;		
XX	pharmaceutical.		
XX	Drosophila melanogaster.		
XX	MO2001171042-A2.		
XX	27-SEP-2001.		
XX	23-MAR-2001; 2001WO-US09231.		
XX	23-MAR-2000; 2000US-191637P.		
XX	10-SEP-2000; 2000US-061415O.		
XX	(PEKE ) PE CORP NY.		

PI Venter JC, Adams M, Li PMD, Myers EM;  
XX MPI: 2001-656860/75.  
DR N-PSDB: ABL02255.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from *Drosophila* and for elucidating cell signalling and cell-cell  
interactions -  
XX  
PS Disclosure: SEQ ID NO 1248; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from *Drosophila*. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
sequences (AB57737-AB572072).  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pt\_sequences.  
XX  
SQ Sequence 704 AA;

Query Match 42.1%; Score 1509.5; DB 22; Length 704;  
Best Local Similarity 44.6%; Pred. No. 2.2e-126;

Matches 311; Conservative 114; Mismatches 230; Indels 43; Gaps 14;

OY 4 YVIGVGRADCTGQVSDINLMGYGKNGQNAAGLTLFLSRAFLIADPDGSNRAAFVSELC 63  
DB 25 YVIGVGRADITGPPEVINEFMGYANIKOVGRGITHVFARAFVDEKGS NRAAFVSADAG 83  
OY 64 MISQRLREVLKRLSKYGSILYRDNVILSAIHTHSGAPGFQYLYLILASGFSNRFFQ 123  
DB 84 MNGYGLKREVIRKRLDARGCNVYHNDVNAISGTHGAPGGLMHLIDYISLIGFQYTFE 143  
OY 124 YVSGIMKSIDIAHNLKPKGKIFINKGVANVQINRSPSSYLLNPSSEARVSNNDKEM 183  
DB 144 VMAQGLYLICIRKATONLDVGRILLSTKYLVANINRSPSSYLRNRAPEAOUEHDTDKL 203  
OY 184 LVKLVDLNGEDLGLISWFAIHPVSMNNSNHFVNSDNMGYAAYLFEDEKKNKCYLPQGGPF 243  
DB 204 TQLRVEDLNNLLGAFNMYAVHATSMNNTNRLVTSNDVGYALLLEKEVNPKNMGKGF 263  
OY 244 VAGFSSNLDGVSPIILGPHCVNTEGSCDNKSTCPNGGSPSCMASGPGQDMFESTHITG 303  
DB 264 VAGFSSNLDGVSPIILGPHCVNTEGSCDNKSTCPNGGSPSCMASGPGQDMFESTHITG 322  
OY 304 RIYYOKA-----KELYASASOEVTGPVLAHOMVNMVDSVOLNAVHT-----VKTC 350  
DB 323 OHLAALAGLNLBOQSESTARREYTDVRFIHOFVMPNPN---GSYTNPLSKVYDKIRKC 379  
OY 351 KPALGYSFPAAGTIDGVSGLINTQGTGEGDFEWDLRDOLLGKPSSEIYECOKPKPILHS 410  
DB 380 QPAMGYSPAAGTTDGPAGFSFEGQTTDPMNMFVRD-FIAAPQEDICHCHEPKPILAT 438  
OY 411 GELLTPHWPQDIYVOIVTGSALAAIPGLITMSGRFREAIKKEKALYGMKDMYV 470  
DB 439 GATPPEWOPKIVSDOLKIGDVIIAVPCETTMAGRRILNQJRAASAAGGIDTEVI 498  
OY 471 IAGLSNVYTHYITTEEYOAOQRYEASTIYGPHTLSAYIOLFRDI AKATATVJNMSSG 530  
DB 499 IAGLNTITSYTVTEEYOAOQRYEASTIYGPHTISYIMDYERLTKAMRN--ETVDAG 556  
OY 531 PEPPFKNLIALSLDNIA-DRAPIGKHGVDLQPAKPEYRGEVEVIFVGANPKNSAEN 589  
DB 557 PSPPMNDVYMLSLNGVLEFDGHPINTDFCYKSDQNKKEGINETKVTYISGNPNNT-- 614  
OY 590 QHOFELTYEK--YEDSVADQIMYNDASWETRFVHK--GILGSLNATIIYHIDTAP 645  
DB 615 FEKTYFTIERKINEDR--MKVAVTASWETKVMHRTNTILGSEMDIYDISPQTL 671

OY 646 GIVRIYFEGHNRKQELLKPAVILA----FEGISSPFEV 679  
DB 672 GEXRIHSGSEYK-----YILGKGYPEGLJHSPTV 701

RESULT 6  
AAV57307  
ID AAV57307 standard; Protein; 646 AA.  
XX  
AC AAV57307;  
XX  
XX

05-JUN-2000 (first entry)

P. aeruginosa mature ceramidase sequence.

Ceramidase; ceramide; sphingosine; fatty acid; atopic dermatitis;  
lipid engineering; antidermatitis.

Pseudomonas aeruginosa.

EP980912-A1.

23-FEB-2000.

20-AUG-1999; 99P-0116154.

20-AUG-1998; 98JP-0234769.

(TAKI) TAKARA SHUZO CO LTD.

Okino N, Ito M;

WPI: 2000-258590/23.

N-PSDB: AA290566.

New Pseudomonas aeruginosa alkaline ceramidase gene, useful for  
diagnosis, treatment and prevention of atopic dermatitis -

Claim 1: Page 16-18; 32pp; English.

XX  
XX This represents a mature alkaline ceramidase from *Pseudomonas aeruginosa*.  
XX Ceramidase can be produced by standard recombinant methodology.  
XX Ceramidase converts ceramide to sphingosine and a fatty acid, causes, or  
exacerbates, atopic dermatitis. The ceramidase gene, polypeptide, derived  
oligonucleotides and ceramidase-specific antibodies are used to prevent  
and/or treat atopic dermatitis. The gene and the polypeptide can be used  
as reagents for lipid engineering and for analyzing the structure and  
activity of ceramides.  
XX  
XX

Sequence 646 AA;

Query Match 30.4%; Score 1090.5; DB 21; Length 646;  
Best Local Similarity 35.9%; Pred. No. 9.8e-89;

Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;

OY 4 YVIGVGRADCTGQVSDINLMGYGKNGQNAAGLTLFLSRAFLIADPDGSNRAAFVSELC 63  
DB 5 YVIGVGRADITGPPEVINEFMGYANIKOVGRGITHVFARAFVDEKGS NRAAFVSADAG 83  
OY 64 MISQRLREVLKRLSKYGSILYRDNVILSAIHTHSGAPGFQYLYLILASGFSNRFFQ 123  
DB 65 MFGQAVHLKVLARLAKAPGVYDENNMVLAATHTHSGPGGFSHYAMVNLVYLGFEKTFN 124  
OY 124 YVSGIMKSIDIAHNLKPKGKIFINKGVANVQINRSPSSYLLNPSSEARVSNNDKEM 183  
DB 125 AIVDGIIVRSIERAQRRLQPGRLYSGGELRNASRRKSLSLHKNP--DIAGYEDGIDPOM 182  
OY 184 LVKLVDLNGEDLGLISWFAIHPVSMNNSNHFVNSDNMGYAAYLFEDEKKNKCYLPQGGPF 243  
DB 183 SVLSFYVDANGELAGLISWFAIHPVSHSTMTNANHILSDNKGYSYHEHYVR-----KSGF 237  
OY 244 VAGFSSNLDGVSPIILGPHCVNTEGSCDNKSTCPNGGSPSCMASGPGQDMFESTHITG 303

```

Db 238 VAAFAQTNMGNSLNL-----NLKPGSGPPDNEFDNIREIG 273
QY 304 RIITYOKAKELLYASASOEVTGPVLAHQWVMTDVSQVLAAT--HTVKTCKPALGYSFAAG 361
Db 274 LRQFAKAYELIAGQAQOEVEVLGELDSRFRFVDFTLPIRPEPTDGPQLCTAAIGTSLIAG 333
QY 362 -TIDVSGLNTIGCTEGDEPFMTLDQLLGKPSSEIVECQKPKILLHSGELTIHPMQ 420
Db 334 STEDGCGPLGLEGG---NNPFSLALGGLTGVPQDELVOCAEKTLIADGNKK-PYPWT 389
QY 421 PDIVDQIYVYVGSLSLAIAIPGELTMSGRFRFAIKKEFALYGMKMTVYIAGLSNVTYH 480
Db 390 PTVLPYQMRIGQLELIGAPAEFTVAGVYIRRAVQAASAAQIRH--VYFNQYANAYAS 447
QY 481 YTTTVEYQAQRYEAASSTIYGPHTLSAYIQLFRLAKA-----IATDVA-----N 526
Db 448 YVTTREYVAAQEEGSGTGLGPTQAAYOOLFPDVAVALERLPEVTSALAPLSCQMN 507
QY 527 MSSGPPPPFFKNLIASLIPNIADRAPIGKHFGDVLQPAKPEYRVGEVVEYIEFGANPKNS 586
Db 508 FQTVG-----VADDPYIGKSGFDVLOOQPRESYRIGDKVTVAFTYGHKKN- 551
QY 587 AENQTHQTELTVEKY-EDSVADWQIMVNDASMETREFYMHKGLISNATYWHIIPDTAYP 645
Db 552 -DLRTEKTELEVYVNIKDKQTPVYATDNDMDTQYRMERVIGISAKATISMSIPGTER 610
QY 646 GIYRIRYFGHNKQKELLKPAVILAFEGISSPEFVVT 682
Db 611 GHYIRHYGNKAKNFWTK---ISEIGSGTSPFEVLGT 644

```

## RESULT 7

AAV57314  
ID AAV57314 standard; Protein; 670 AA.

XX AAV57314;

DT 05-JUN-2000 (first entry)

XX P. aeruginosa ceramidase sequence.

XX Ceramidase; ceramide; sphingosine; fatty acid; atopic dermatitis;

KW lipid engineering; antidermatitis.

XX Pseudomonas aeruginosa.

XX BP980912-A1.

XX 23-FEB-2000.

XX 20-AUG-1999; 99EP-0116154.

XX 20-AUG-1998; 98JP-0234769.

XX (TAKI ) TAKARA SHUZO CO LTD.

XX Okino N, Ito M;

XX MPI: 2000-258590/23.

XX N-PSDB; AA290573.

XX New Pseudomonas aeruginosa alkaline ceramidase gene, useful for

XX diagnosis, treatment and prevention of atopic dermatitis -

XX Example 1; Page 24-27; 32pp; English.

XX This represents a ceramidase from Pseudomonas aeruginosa strain AN-17.

XX Ceramidase can be produced by standard recombinant methodology.

XX Ceramidase converts ceramide to sphingosine and a fatty acid, causes, or

XX exacerbates, atopic dermatitis. The ceramidase gene, polypeptide, derived

XX oligonucleotides and ceramidase-specific antibodies are used to prevent

XX and/or treat atopic dermatitis. The gene and the polypeptide can be used

XX as reagents for lipid engineering and for analyzing the structure and

CC activity of ceramidases.

XX Sequence 670 AA;

Query Match 30.4%; Score 1090.5; DB 21; Length 670;  
Best Local Similarity 35.9%; Pred. No. 1e-88;  
Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;

```

QY 4 YVIGVRADCTGVSDINIMGYKNGCOMAGLLTFLRFAFTIADPDGNSRMAFVSVELC 63
Db 29 YRGLGKADITGGAEEVGMWGSLSBQKTAGIHMRQMARAFVIEEASGRLLYVNTDGL 88
QY 64 MISQRLRELYKLKESKYSGLYRDVYILSAITHHSGPAGFPQYTYLILASSEFSKRTQ 123
Db 89 MTFQAVHLKYLAKKAPGVYDENVNMLAATHTHSGPGGFSHYAYNYLSVLGFEKTEEN 148
QY 124 YVSGIMKSIDIAHTNLKPKIFINKGNVANVOINRSPSSYLNPQSEARYSNDKEM 183
Db 149 AIVDGIYRSTERRQARLQPRRLFYGGSELRLNBSRNSLSHLKNP--DIAGYEDGIDPQ 206
QY 184 LVKLVDLNGEDLGLISWFAIHPVSNMNSNHFVNSDNMGYAAYLFEQKKNKGYLPQGP 243
Db 207 SVLSFDANGELAGAISWFEVHSTSMTNANHLISPONKGYASVHMEHDVSR-----KSG 261
QY 244 VAGFASSNLDGVPNLIHPCHVNTGSCDNDKCTCNGGSMKMGAGPGQDMFEETHIG 303
Db 262 VAAFAQTNMGNSLNL-----NLKPGSGPPDNEFDNIREIG 297
QY 304 RIITYOKAKELLYASASOEVTGPVLAHQWVMTDVSQVLAAT--HTVKTCKPALGYSFAAG 361
Db 298 LRQFAKAYELIAGQAQOEVEVLGELDSRFRFVDFTLPIRPEPTDGPQLCTAAIGTSLIAG 357
QY 362 -TIDVSGLNTIGCTEGDEPFMTLDQLLGKPSSEIVECQKPKILLHSGELTIHPMQ 420
Db 358 STEDGCGPLGLEGG---NNPFSLALGGLTGVPQDELVOCAEKTLIADGNKK-PYPWT 413
QY 421 PDIVDQIYVYVGSLSLAIAIPGELTMSGRFRFAIKKEFALYGMKMTVYIAGLSNVTYH 480
Db 414 PTVLPYQMRIGQLELIGAPAEFTVAGVYIRRAVQAASAAQIRH--VYFNQYANAYAS 471
QY 481 YTTTVEYQAQRYEAASSTIYGPHTLSAYIQLFRLAKA-----IATDVA-----N 526
Db 472 YVTTREYVAAQEEGSGTGLGPTQAAYOOLFPDVAVALERLPEVTSALAPLSCQMN 531
QY 527 MSSGPPPPFFKNLIASLIPNIADRAPIGKHFGDVLQPAKPEYRVGEVVEYIEFGANPKNS 586
Db 532 FQTVG-----VADDPYIGKSGFDVLOOQPRESYRIGDKVTVAFTYGHKKN- 575
QY 587 AENQTHQTELTVEKY-EDSVADWQIMVNDASMETREFYMHKGLISNATYWHIIPDTAYP 645
Db 576 -DLRTEKTELEVYVNIKDKQTPVYATDNDMDTQYRMERVIGISAKATISMSIPGTER 634
QY 646 GIYRIRYFGHNKQKELLKPAVILAFEGISSPEFVVT 682
Db 635 GHYIRHYGNKAKNFWTK---ISEIGSGTSPFEVLGT 668

```

## RESULT 8

AAV38464  
ID AAV38464 standard; Protein; 280 AA.

XX AAV38464;

XX 18-OCT-1999 (first entry)

XX Human secreted protein encoded by gene No. 21.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;

XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;

XX developmental abnormality; foetal deficiency; blood; allergy; renal;

XX immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;

XX inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;

XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS	Homo sapiens.
XX	MO935158-A1.
XX	15-JUL-1999.
XX	06-JAN-1999; 99WO-US00108.
XX	07-JAN-1998; 98US-0070704.
XX	07-JAN-1998; 98US-0070657.
XX	07-JAN-1998; 98US-0070656.
XX	07-JAN-1998; 98US-0070692.
XX	(HUMA-) HUMAN GENOME SCI INC.
XX	Brewer LA, Duan RD, Edner R, Lafleur DW, Ni J;
XX	Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX	WPI: 1999-444190/37.
XX	N-PSDB; AA206239.
XX	New isolated human genes and the secreted polypeptides they encode
XX	Disclosure: Page 213-214; 227pp; English.
XX	This sequence represents a secreted human protein encoded by the gene
XX	clone detailed in the descriptor line. The gene can be used to generate
XX	fusion proteins by linking to the gene to a human immunoglobulin Fc
XX	portion (e.g. AA206210) for increasing the stability of the fused
XX	protein as compared to the human protein only.
XX	The invention relates to 36 novel genes and their fragments (nucleic
XX	acid sequences: AA206219-206263; amino acid sequences: AA238386-238498)
XX	which are useful for preventing, treating or ameliorating medical
XX	conditions e.g. by protein or gene therapy. Also, pathological
XX	conditions can be diagnosed by determining the amount of the new
XX	polypeptides in a sample or by determining the presence of mutations in
XX	the new polynucleotides. Specific uses are described for each of the 36
XX	polynucleotides, based on which tissues they are most highly expressed in
XX	(see AA206219 for described uses).
XX	Sequence 280 AA:
XX	Query Match 24.5%; Score 879; DB 20; Length 280;
XX	Best Local Similarity 81.1%; Pred. No. 2.6e-70;
XX	Matches 163; Conservative 28; Mismatches 10; Indels 0; Gaps 0
QY	1 FSGYIGVGRADCTGOVSDIMMGYGGKGNARGLTFLFRFAFLADPDGSGNRMAFVSV 60
DB	61 FSGYHIGVGRADCTGOVADINLMGKSGGNQAAGLTFLRYSRAFLMAEPDGSNRIVFSI 120
YY	61 ELCMISQRLRLVETLKRLESKGSLYRDNVILSAITHHSGPAGFYQYLYLIIASGEFSNR 120
DB	121 DIGMSQRLRLVETLKRLESKGSLYRDNVILSGTHHSGPAGFYQYLYLIIASGEFSNQ 180
QY	121 TFOYIVSGIMKSIDIAHTNMLEPKGIFIKKGVANVOINRSPSYLLNQSEFARRSSMTD 180
DB	181 TFOHMYVIGILKSIDIAHTNMKPGKIFINKGVNDGVOINRSPSYLLNQSEFARRSSMTD 240
QY	181 KEMVLVKLVNDGDELGLISW 201
DB	241 KEMVLVKLVNDGDELGLISF 261
XX	RESULT 9
XX	ID AAB23290 standard; peptide: 35 AA.
XX	AC AAB23290;
XX	02-FEB-2001 (first entry)

```

DE Mouse neutral/alkaline ceramidase peptide, SEQ ID NO:3.
XX
XX Neutral/alkaline ceramidase: mouse; murine; recombinant production;
KW cellular ceramide content control; antibody; membrane synthesis;
KW lipid engineering; ceramide metabolism; drug development;
KW mature ceramidase;
XX
XX Mus sp.
OS
XX WO200058448-A1.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000; 2000WO-JP01802.
XX
XX 26-MAR-1999; 99JP-0084743.
XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX Ito M;
XX
XX WPI: 2000-619079/59.
XX
XX Mammalian neutral/alkaline ceramidase applicable as lipid engineering
XX reagent for studying structure and functions of ceramide as well as
XX developing drugs for diseases associated with ceramide metabolism
XX
XX Example 2; Page 51; 76pp; Japanese.
XX
XX The invention relates to a mouse neutral/alkaline ceramidase and to
XX nucleic acids encoding it. The invention also relates to expression
XX vectors and host cells comprising DNA encoding mouse neutral/alkaline
XX ceramidase, the recombinant production of the ceramidase, an antibody
XX against the ceramidase, and a method for controlling the ceramide
XX content in cells and/or tissues by introducing the gene or its antisense
XX nucleic acid into the cells and/or tissues. The ceramidase can be used
XX as a lipid engineering reagent for studying structure and functions of
XX ceramide as well as developing drugs for diseases associated with
XX ceramide metabolism. The present sequence represents the mature mouse
XX neutral/alkaline ceramidase.
XX
XX Sequence 35 AA:
XX
XX Query Match 4.7%; Score 167; DB 21; Length 35;
XX Best Local Similarity 91.4%; Pred. No. 9.1e-08;
XX Matches 32; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 235 GYLPGGPFFVAGFASSNLGDVSPNIIPLPHCVNIGE 269
XX ||||||| ||||||| ||||||| |||||
XX 1 GYLPGGPFFVNGFASSNLGDVSPNIIPLPXVNTGE 35
XX
XX RESULT 10
XX ID AAG03568
XX AC AAG03568 standard; Protein: 80 AA.
XX AC AAG03568;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein, SEQ ID NO: 7649.
XX
XX Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.
XX
XX Homo sapiens.
XX
XX EPI033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX

```

XX 26-FEB-1999; 9905-0122487.  
PR (GEST ) GENSET.  
XX  
XX  
PI Dumas Mline Edwards J, Duclert A, Giordano J;  
XX WPI: 2000-500381/45.  
DR N-PSDB: AAC03574.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 13; SEQ ID 7649; 71pp + CD-ROM; English.  
XX  
XX The present sequence is a polypeptide encoded by one of a large number  
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
CC different tissues. EST sequences usually correspond mainly to the 3'  
CC untranslated region (UTR) of the mRNA because they are often obtained  
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.  
XX  
XX Sequence 80 AA:  
SO  
Query Match 4.1%; Score 148; DB 21; Length 80;  
Best Local Similarity 35.7%; Pred. No. 1.9e-05;  
Matches 35; Conservative 9; Mismatches 22; Indels 32; Gaps 2;  
OY 479 THYITTYEYQKREASTIGPHTLSAIFQFLDLAKAINTDVANNSSGPE---PP 534  
DB 11 THYLTFSS-----SEKLAAGYCYGCHSQPEQRSTRS 42  
OY 535 FFKNLASLIPNIADRPAGKHFQDLPAPKPEYRGE 572  
DB 43 LEQTIIVPLPSIVDRAPKGRFEGDVLQAPKPEYRGE 80  
RESULT 11  
ID AAM18305  
AAW18305 standard; Protein: 1565 AA.  
XX  
XX AAM18305:  
AC  
XX  
XX 30-JAN-1998 (first entry)  
DT  
XX  
XX Photornhabdus luminescens insect toxin TcCB.  
DE  
XX  
XX Insecticide; insect; toxin; pest control; biological control;  
KM Photornhabdus luminescens; TcCB; Southern corn rootworm;  
KM Colored potato beetle; Western corn rootworm; meal worm;  
KM boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
KM cabbage looper; codling moth; corn earworm; European corn borer;  
KM tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
KM Diptera; Dictyoptera; Acarina; Homoptera.  
XX  
XX Photornhabdus luminescens strain W-14 (ATCC 55397).  
OS  
XX  
XX Key Location/Qualifiers  
FH 1.11  
FT Peptide /note="N-terminal peptide (Claim 30)"  
XX  
XX WO9717432-A1.  
XX  
XX 15-MAY-1997.  
PD  
XX

PF 06-NOV-1996; 96MO-US18003.  
XX  
XX 28-AUG-1996; 9605-0705484.  
PR 06-NOV-1995; 9505-0007255.  
PR 28-FEB-1996; 9605-0608423.  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA;  
XX  
XX WPI: 1997-281022/25.  
DR N-PSDB: AAT68849.  
XX  
XX Photornhabdus sp. insecticidal protein toxins and DNA encoding them -  
PT can be genetically engineered into insect larvae food and plants for  
PT insect control  
XX  
XX Claim 34; Page 239-245; 276pp; English.  
XX  
XX This polypeptide comprises the 185 kDa TcCB insecticidal toxin  
CC protein of Photornhabdus luminescens W-14. Its sequence was deduced  
CC from a genomic DNA clone (AAT68849) and includes an isolated  
CC N-terminal peptide (see AAM17867). Claimed toxins of P. luminescens  
CC (see AAM17871, AAM17884-89, AAM17899-900, AAM18301-06) can be produced  
CC by recombinant DNA methods and applied to, or genetically engineered  
CC into, insect larvae food and plants for insect control. The toxins  
CC are particularly effective against Southern corn rootworm, Colorado  
CC potato beetle, Western corn rootworm, meal worm, boll weevil and  
CC turf grub (Coleoptera), beet armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer, tobacco hornworm  
CC and tobacco budworm (Lepidoptera), and are also active against  
CC insects of the orders Hymenoptera, Diptera, Dictyoptera, Acarina  
CC and Homoptera: (All claimed).  
XX  
XX Sequence 1565 AA:  
SO  
Query Match 4.1%; Score 146.5; DB 18; Length 1565;  
Best Local Similarity 20.9%; Pred. No. 0.0039;  
Matches 144; Conservative 78; Mismatches 253; Indels 215; Gaps 33;  
OY 37 TRLFSRAFLADPDGSKRMATVSELCMISQRLRLKLESKYGLSRDNYLSAII 96  
DB 656 TYTLSEADPSTDPDXN-----YLQVCL-----NYWDHYD 685  
OY 97 THSGPAGFQ-----YTLYLASEGFSNTPQYIYS-----GIMKSIDIAHTNLKPKKI 145  
DB 686 RPSGKKGATSWVSKMFNVYVALQDSKAPDAIRLVSRTDSKRGLOYIDFWTSSL-PAKT 744  
OY 146 FINKGNV-----ANVOINRSPSSYLL--NPOSERARSNTDKMLYLKLVLDNGEDLG 197  
DB 745 RLNTTFVRLILKANIIGLD-SLDDYTLQADPLEADLYTDGKSEPM-----DFNGSN-G 796  
OY 198 LISMFALHPVSMNNSNHFVNSDNMGYAA-----YLFQE-KMKGYLPQGGPYPAGRASSNL 252  
DB 797 LYFWELFEHLPLVATRFANEOQFSPAOKSLHYIDPAMKNN---PHNAP-----AY 845  
OY 253 GVSFNIILGPHVNVNAGESDNDKSTCPNCGSPSCMASAGGQGMFESTHIGRIIYOKA-- 310  
DB 846 WNVRLPVEG---NSDLSRHLDSDIDP-----DTQAYAH---PVIYQKAYF 884  
OY 311 -----KELYASASQEVYTG-----VLAHH 329  
DB 885 IAYVSNLIAQGDMMYROLTRDGLTQARVYYNLAELGGRPDVLSSTWPTQTLDTLAAG 944  
OY 330 QWVNMNTDVSVOLNATHYKTKCPALGYSPACTIDGV---SGANTIQGTGEGPFDYLRD 387  
DB 945 OKAVLRLDFEHOJLANSDTALPALGRNVSYLKLADNGYFNEPLNVLMS-----HMDTDLA 999  
OY 388 QLL-----GKPSSEIVECQKPKPIL-----HSGELTTPHMQPDIVVOIVTJGS 433

DB 1000 RLYNLRHNLTVDGKPLSLPYAPVDPVALLAQAQSGTLT-----M:VSGAMLTVGT 1052  
OY 434 L-AIAIP-----GELTTMSG-----RRFRDAIKKEPALYGMKDMT-----VYIA 472  
DB 1053 LTFSPMLPRAYSAVGLTISFGONLSLSESRACOEELAQOOLDMSSYATLQOQALD 1112  
OY 473 GLSNVYTHYITTYEEQAOQRYEAASITYGPHLSAVITQLEFRLAKA-----IATDYYANMS 528  
DB 1113 GLAADRLALLASQAQTAQ-QRHDHYTLY-QNNISSAEQOLVMTQTSASGLISSSTGVQTA 1170  
OY 529 SCPEPPEFFKNLILASLPINADRAPIGKHFQDVLQPAKPEYRGEVEVYFVGANKPNSAE 588  
DB 1171 SG-----ALKVIPNIFGLADGGS-----RREGVTEAIAIGMAAGQAT 1208  
OY 589 NOTHOTFLTEKYEEDSVADMOIMYNDASWE 618  
DB 1209 SYVAERLATENTYRRRREEMOIOYQAOSE 1238  
RESULT 12  
AAM56568  
ID AAM56568 standard; Protein: 1565 AA.  
XX AAM56568;  
AC AAM56568;  
XX AAM56568;  
DT 07-AUG-1998 (first entry)  
DE Toxin TccB, encoded by the tccB gene from genomic region tcc.  
XX  
XX Photorhabdus luminescens W-14; nematode; symbiotic;  
KW Heterorhabditis; tca; tcb; tcd; insecticidal activity; toxin;  
KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
KW Homoptera; Southern corn rootworm; Colorado potato beetle;  
KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW tobacco hornworm; budworm.  
XX  
XX Photorhabdus luminescens.  
OS  
XX  
XX W09808932-A1.  
PN  
XX  
XX 05-MAR-1998.  
PD  
XX  
XX 05-MAY-1997; 97WO-US07657.  
PF  
XX  
XX 06-NOV-1996; 96WO-US18003.  
PR  
XX 28-AUG-1996; 96US-0705484.  
PR 06-NOV-1996; 96US-0743699.  
XX  
XX (DOMC) WISCONSIN ALUMNI RES FOUND.  
PA  
XX  
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fattig R;  
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA, Sukhinda K;  
XX  
XX MPI: 1998-179427/16.  
DR N-PSDB; AAV29926.  
XX  
XX Isolated toxins from Photorhabdus luminescens strains - useful for  
PT control of insect pests  
XX  
XX  
PS Claim 34; Pages 276-281; 321pp; English.  
XX  
XX The present sequence represents a protein named TccB of the bacterium  
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
CC produced from these regions that are associated with insecticidal  
CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,

CC baculoviruses or microbial hosts for toxin production. They can be used  
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.  
XX  
XX  
SQ Sequence 1565 AA.  
Query Match 3.7%; Score 131.5; DB 19; Length 1565;  
Best Local Similarity 20.4%; Pred. No. 0.086;  
Matches 141; Conservative 77; Mismatches 257; Indels 215; Gaps 32;  
OY 37 TLFSRAFLPADPGSNRAAFVSEICMISQRLREVLKRLSKYGSGLRRNVLISAH 96  
DB 656 YTTLSADFTDPDKN-----YLOYCL-----NVDWHDYD 685  
OY 97 THSGPAGPEQ-----YTLXILASEGFSNRPQYIVS-----GIMKSIDIAHTNLKPKRT 145  
DB 686 RPSGKKGAYSWSKMNENYVALQDSKAPPAIRIVSRYSKGLGYDLFWTSLSL-PAKT 744  
OY 146 FTKGNV-----ANYQIRSPSSYLL--NPSERARVSSNTDKMLYLKLVLDNGEDIG 197  
DB 745 RLNTFVFRILIEKANIGLD-SLDLYTLQADPSLEADLVTDGKSEPW-----DFNSN-G 796  
OY 198 LISWFAIHPSVNNNSNHFVNSDNMGYAA-----YLFQGE-KNKGILPGGSPFVAGFSSNL 252  
DB 797 LYFWLEFHLPLVLAIRFANECQFSPQAKSLHYTFDPANKN-PPHNP-----AY 845  
OY 253 GDSVPIILPHCVNTGESCNDKSNCPNGSPGCMASGCGQMFESTHITGRITTYOKA-- 310  
DB 846 WNVRLPVEG-----NSDLSHLDSDIDP-----DTQAYAH--PYIQKAYF 884  
OY 311 -----KELYASASQEVYGP-----VLAH 329  
DB 885 IAYVSNLIAQDMWYRQLTRDGLTQARVYNNLAELGPRPVLSISWTPTQTLDTLAAG 944  
OY 330 QWVNMTPDVSVQNAHTFVATCKPALGYSFAACTIGV--SGNITQGTGEQPFMDTLND 387  
DB 945 OKAVLDFEHOANSSTALPALPGRNVSTLKLADNGEYNEPLNMLS-----HMDTLDA 999  
OY 388 QLL-----GKPSSEIVECOKPKPIL-----HSGELTTPHPWQPDIVQVITVGS 433  
DB 1000 RLYNLRHNLTVDGKPLSLPYAPVDPVALLAQAQSGTLT-----NGVSGAMLTVP 1052  
OY 434 LAIAI-----PGEITMSG-----RRFRDAIKKEPALYGMKDMT-----VYIA 472  
DB 1053 YRFSAMLPRAYSAVGLTISFGONLSLSESRACOEELAQOOLDMSSYATLQOQALD 1112  
OY 473 GLSNVYTHYITTYEEQAOQRYEAASITYGPHLSAVITQLEFRLAKA-----IATDYYANMS 528  
DB 1113 GLAADRLALLASQAQTAQ-QRHDHYTLY-QNNISSAEQOLVMTQTSASGLISSSTGVQTA 1170  
OY 529 SCPEPPEFFKNLILASLPINADRAPIGKHFQDVLQPAKPEYRGEVEVYFVGANKPNSAE 588  
DB 1171 SG-----ALKVIPNIFGLADGGS-----RREGVTEAIAIGMAAGQAT 1208  
OY 589 NOTHOTFLTEKYEEDSVADMOIMYNDASWE 618  
DB 1209 SYVAERLATENTYRRRREEMOIOYQAOSE 1238  
RESULT 13  
AAY38468  
ID AAY38468 standard; Protein: 27 AA.  
XX AAY38468;  
AC AAY38468;  
XX AAY38468;  
DT 18-OCT-1999 (first entry)  
XX  
XX Human secreted protein encoded by gene No. 21.  
DE  
XX

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
OS Homo sapiens.  
XX  
PN WO935158-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 06-JAN-1999; 99WO-US00108.  
XX  
PR 07-JAN-1998; 98US-0070704.  
PR 07-JAN-1998; 98US-0070657.  
PR 07-JAN-1998; 98US-0070658.  
PR 07-JAN-1998; 98US-0070692.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;  
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;  
XX  
DR WPI: 1999-444190/37.  
DR N-PSDB: AA206239.  
XX  
PT New isolated human genes and the secreted polypeptides they encode  
XX  
PS Disclosure; Page 215; 227pp; English.  
XX  
SS This sequence represents a secreted human protein encoded by the gene  
CC clone detailed in the descriptor line. The gene can be used to generate  
CC fusion proteins by linking to the gene to a human immunoglobulin Fc  
CC portion (e.g. AA206210) for increasing the stability of the fused  
CC protein as compared to the human protein only.  
CC The invention relates to 36 novel genes and their fragments (nucleic  
CC acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 36  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AA206219 for described uses).  
XX  
SQ Sequence 27 AA:  
XX  
OY Query Match 3.5%; Score 127; DB 20; Length 27;  
Best Local Similarity 88.9%; Pred. No. 0.00023;  
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Db 76 RLESKYGSLYRRNVLTSATHTSGPA 102  
1 RLOSKYGSLYRRNVLTSATHTSGPA 27  
XX  
RESULT 14  
AAY38466  
ID AAY38466 standard; Protein; 26 AA.  
XX  
AC AAY38466;  
XX  
DT 18-OCT-1999 (first entry)  
XX  
DE Human secreted protein encoded by gene No. 21.  
XX  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
OS Homo sapiens.  
XX  
PN WO935158-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 06-JAN-1999; 99WO-US00108.  
XX  
PR 07-JAN-1998; 98US-0070704.  
PR 07-JAN-1998; 98US-0070657.  
PR 07-JAN-1998; 98US-0070658.  
PR 07-JAN-1998; 98US-0070692.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;  
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;  
XX  
DR WPI: 1999-444190/37.  
DR N-PSDB: AA206239.  
XX  
PT New isolated human genes and the secreted polypeptides they encode  
XX  
PS Disclosure; Page 214; 227pp; English.  
XX  
SS This sequence represents a secreted human protein encoded by the gene  
CC clone detailed in the descriptor line. The gene can be used to generate  
CC fusion proteins by linking to the gene to a human immunoglobulin Fc  
CC portion (e.g. AA206210) for increasing the stability of the fused  
CC protein as compared to the human protein only.  
CC The invention relates to 36 novel genes and their fragments (nucleic  
CC acid sequences: AA206219-206263; amino acid sequences AAY38386-Y38498)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 36  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AA206219 for described uses).  
XX  
SQ Sequence 26 AA:  
XX  
OY Query Match 3.5%; Score 125; DB 20; Length 26;  
Best Local Similarity 84.6%; Pred. No. 0.00033;  
Matches 22; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Db 10 RADCTGVSDINIMGYGKNGQNRGL 35  
1 RADCTGVADINIMGYGKSGQNRGI 26  
XX  
RESULT 15  
ABB47751  
ID ABB47751 standard; Protein; 2013 AA.  
XX  
AC ABB47751;  
XX  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes protein #455.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX  
OS Listeria monocytogenes.  
XX  
PN WO20017735-A2.



XX 18-Oct-2001.  
PD  
XX  
XX 11-APR-2001; 2001MO-FR01118.  
PF  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
XX  
PA (INSP ) INST PASTEUR.  
XX  
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi B, Dehoux P,  
PI Dussuguet O, Chectouani F, Nedjari H, Glaser P, Kunst F, Cossart P  
PI Daniels J, Goebel W, Krefz J, Kuh M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L,  
PI Perez-Diaz J, Baquero E, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maugueno E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
DR WPI; 2002-010914/01.  
XX  
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and  
PT related polypeptides -  
XX  
XX Claim 6; SEQ ID No 456; 192pp; French.

The present invention relates to the genome sequence of *Listeria monocytogenes* ECD-e (see ABA03041). The genome sequence and fragments of it are useful for selecting probes and primers for detecting genes in *L. monocytogenes* and related organisms, and for studying genetic polymorphisms and other genomes. The present sequence is a protein encoded by the genome sequence of the present invention. Proteins expressed from the genome sequence are useful for raising specific antibodies. Identification of *L. monocytogenes* and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate *L. monocytogenes*-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccines compositions for the treatment or prevention of infections by *L. monocytogenes* and related organisms.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://ftp.wipo.int/pub/published_pat_sequences).

sq Sequence 2013 AA;

Query Match	3.3%	Score 118.5	DB 23	Length 2013
Best Local Similarity	19.8%	Pred. No. 2		
Matches 148	Conservative 88	Mismatches 266	Indels 247	Gaps 38

```

QY      4  FILADPGSGRRMAF-----VSEVLCMISORLEVLTKRESYSGLYR--RDVNLISAI  95
Db      102 YLLASQTKYELEKRGDKYQAKDLILITKNVDEEFAKVEKSNRQLRSSISDKLFLQA-  166
QY      96 HTHSGPAGFQYLYLLASEGFSNRTFOYVSGIMKSIDIAHTNLPKPIFINKGNVAV  155
Db      161 --DSKATLANYTEQI-----TFNYSINFLDGS-----TINKKGLVIDNN--SNL  203
QY      156 QIMRSPSYLLINQSRARYSSNTDEKEMLYKLV--LNGE--DLGLI--SWEAIHPYSV  209
Db      204 ELVNYPRDPAANNINILTSYSALTQK--LTIMLVNISGAPFPIPIYVRAGYAKGQGLP  261
QY      210 NNSNHFPVNSDNNMGYAY-----LFEQEKKNGYLPGGPFVAG-----  246
Db      262 MNLKAILSGENSSGGTYTPSEKTTVTYNLEESSNDY----SPTTAGNSMAFNKLKELSY  317
QY      247 -----FASSMLGDV--SPNILGPHCVNTGE-----  268
Db      318 SLKPGGYTIQWPEIQKSLKLENSPFNKLLEYLKEKGQGVIVSNADPVIINFGEPFWSQL  377
QY      270 SCNDNR-STCPNPGSPMCASGPGQGDMESTHIIICRILLYQAKELVYAS-----ASQEV  321

```

```

Db 378 STYNGKANVYNLNDKQVVEYEGP-----INANTYKQIKQYSMAAKIPADVBETEX 4427
QY 322 TGPVLAHOMVNMNTDVSVOJANATHYVKTCPALGYSFAAGTIDG-VSGJNITQGTTEGD 3800
Db 428 TGTAVNYDDEVFITTSIKI-----KAEVADASATSIANDSKVSKTSTI----SEGD1 4722
QY 381 F-WDTLEDDOLLKPSSEIVECOKPKP-----TLLHSGELT1PH 418
Db 473 FEMGEMPRVSSAAPGVNDEIAPVPEGIKALSYIPNNNSMASTIKLEYQNGKWSMAP 5322
QY 419 -----WQPIVDVQIYTVGSLAIAPG-----ELTTSGCRPREAI 4555
Db 533 QTSSEGWDFSKIDQSVNRIEKLTFRSDGIINDKDPYTHGTIRMQNIGVAGSEF--TL 5900
QY 456 KKEFALGKMDMT-VVIAGLSNVYHYITTEYEQAQYEAASITTYGPHLTSAI----- 5099
Db 591 RPESITTYTPDKASKTIDTTNNSYKKNQVVE-----KTSAPAKINEVFLSPSTAGTSKG 645
QY 510 -----QJERDLAKAIATDVANMSSGDEPEFFENLASLIPIMABAPKGFQGVOLPAK 5655
Db 646 GFGSTTFEFGDKIAQSVNRGLSYGSKLENNY-----TEVVAPGIDLETM-RNF--IQP-- 6966
QY 566 PEYRAGEVEVIVGNN-----PKNSAENQHOTFLVEKEDSVADWQIMYNDASMET 6129
Db 697 --YR--STLWYTYAPANGTNTLYXPSSAP-----YKGETLSDGSTLY----- 7355
QY 620 RPYWHKGLGLSNATYWHIIPDTAY-PGI 647
Db 736 -----YWEAPDTGLAPGM 748

```

Search completed: July 3, 2003, 12:45:02  
Job time : 76 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 12:39:23 ; Search time 25 Seconds

(Without alignments)  
1131.474 Million cell updates/sec

Title: US-09-937-521-14

Perfect score: 3582  
Sequence: 1 FSGYIGVGRADCTGVSDI.....KPAVILAFGSISSFEEVTT 682

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	3.2	1331	1	CTA2_HUMAN
2	109	3.0	752	1	G9uhc6 homo sapien
3	109	3.0	1953	1	F06681 homo sapien
4	106.5	3.0	665	1	B1GA_SALTY
5	106.5	3.0	2628	1	F25927 salmonella
6	106	3.0	1376	1	FLID_BORBU
7	105.5	2.9	440	1	051845 porphyromon
8	104	2.9	1330	1	CRBH_HUMAN
9	103.5	2.9	557	1	P23445 bacillus su
10	103	2.9	451	1	P22985 rattus norv
11	103	2.9	1012	1	P50504 debaryomyce
12	102.5	2.9	582	1	ARP2_PLAFA
13	102.5	2.9	1026	1	P08364 plasmodium
14	102.5	2.9	1027	1	P09864 avian infec
15	101.5	2.8	754	1	HS60_SCHPO
16	101.5	2.8	4725	1	P25159 drosophila
17	101	2.8	1399	1	P55197 homo sapien
18	100.5	2.8	1715	1	09pn94 campylobact
19	100	2.8	379	1	P43036 dictyosteli
20	99.5	2.8	669	1	09hwc9 pseudomonas
21	99.5	2.8	1712	1	063374 rattus norv
22	98.5	2.7	836	1	P94891 leptospira
23	98	2.7	864	1	P54804 canis fami1
24	98	2.7	868	1	O9P282 homo sapien
25	97.5	2.7	376	1	P47280 mycoplasma
26	97.5	2.7	882	1	O03351 rattus norv
27	96.5	2.7	510	1	P43158 porphyromon
28	96.5	2.7	1292	1	P38501 alcaligenes
29	96.5	2.7	2051	1	P38501 alcaligenes
30	96	2.7	637	1	P49961 homo sapien
31	96	2.7	1012	1	P47582 mycoplasma
32	96	2.7	1012	1	P07149 s fatty aci
33	95.5	2.7	3130	1	O05961 bruceella ov
					P15480 avian infec
					P22351 avian infec
					P06073 homo sapien

34	95	2.7	484	1	YE51_METUA	O58846 methanococc
35	95	2.7	553	1	H7U1_TREVO	O97824 thermoplasm
36	94.5	2.6	240	1	GTOI_MOUSE	O09131 mus musculum
37	94.5	2.6	331	1	DIV_ECOLI	P15286 escherichia
38	94.5	2.6	518	1	GLYM_SOLTU	P50433 solanum tub
39	94	2.6	431	1	GLAI_RHIME	O92946 rhizobium m
40	94	2.6	501	1	DLDP_PEA	P31023 pismum sativ
41	94	2.6	501	1	GSPE_AERHY	P1741 aeromonas h
42	94	2.6	593	1	DCPI_PEA	P51850 pismum sativ
43	94	2.6	706	1	C1AA_PARP	O45358 paenibacill
44	94	2.6	1447	1	DCG_HUMAN	P43146 homo sapien
45	94	2.6	2415	1	PGCA_HUMAN	P16112 homo sapien

#### ALIGNMENTS

RESULT 1  
CTA2\_HUMAN STANDARD; PRT; 1331 AA.  
ID CTA2\_HUMAN  
AC G9uhc6; G9UQ12; 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Contactin associated protein-like 2 precursor (Cell recognition molecule Caspr2).  
CN CNTNAP2 OR CASPR2 OR KIA0868.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
RC TISSUE=Brain;  
RX MEDLINE=20088303; PubMed=10624965;  
RA Poliak S., Gollan L., Martinez R., Custer A., Einheber S.,  
RA Salzer J.L., Trimmer J.S., Shrager P., Peles E.;  
RT "Caspr2, a new member of the neuroxin superfamily, is localized at the  
RT juxtaparanodes of myelinated axons and associates with K<sup>+</sup> channels.";  
RL Neuron 24:1037-1047(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=2125095; PubMed=11352571;  
RA Nakabayashi K., Scherer S.W.;  
RT "The human contactin-associated protein-like 2 gene (CNTNAP2) spans  
RT over 2 Mb of DNA at chromosome 7q35.";  
RL Genomics 73:108-112(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99156230; PubMed=10048485;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:355-364(1998).  
RN [4]  
RP FUNCTION: MAY PLAY A ROLE IN THE FORMATION OF FUNCTIONAL DISTINCT  
RP DOMAINS CRITICAL FOR SALVATORY CONDUCTION OF NERVE IMPULSES IN  
RP MYELINATED NERVE FIBERS. SEEMS TO DEMARCAE THE JUXTAPARANODAL  
RP REGION OF THE AXO-GLIAL JUNCTION.  
CC -1- SUBUNIT: ASSOCIATES WITH KCNA2 (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).  
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN NERVOUS SYSTEM.  
CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.  
CC -1- SIMILARITY: CONTAINS 4 LAMININ G-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF193613; AAF25199.1; -  
 DR EMBL: AF319045; AAK34932.1; -  
 DR EMBL: AF318292; AAK49902.1; -  
 DR EMBL: AF318298; AAK49903.1; -  
 DR EMBL: AB020675; BAA74891.1; ALT\_INIT.  
 DR HSSP: P12259; ICZT.  
 DR Genew: HGNC:13830; CNTNAP2.  
 DR MIM: 604569; -  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR00421; FA58\_C.  
 DR InterPro: IPR01092; HLH\_basic.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR003585; Neurexin-like.  
 DR Pfam: PF00008; EGF\_4.  
 DR Pfam: PF00054; Laminin\_G; 6.  
 DR Pfam: PF00754; F5\_F8\_type\_C; 2.  
 DR SMART: SM00294; EGF\_1.  
 DR SMART: SM00181; EGF; 1.  
 DR SMART: SM00231; FA58C; 1.  
 DR SMART: SM00282; Lame; 4.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE: PS01285; FA58C\_1; 1.  
 DR PROSITE: PS01286; FA58C\_2; 1.  
 DR PROSITE: PS50025; LAM\_G\_DOMAIN; 4.  
 DR GlycoProtein: Cell adhesion; Repeat.  
 DR SIGNAL: 1  
 DR CHAIN: 27  
 DR DOMAIN: 28 1331  
 DR TRANSMEM: 1263 1283  
 DR DOMAIN: 1284 1331  
 DR DOMAIN: 38 178  
 DR DOMAIN: 216 368  
 DR DOMAIN: 401 552  
 DR DOMAIN: 554 591  
 DR DOMAIN: 837 963  
 DR DOMAIN: 963 1002  
 DR DOMAIN: 1055 1214  
 DR CAROHD: 289 289  
 DR CAROHD: 346 346  
 DR CAROHD: 363 363  
 DR CAROHD: 379 379  
 DR CAROHD: 436 436  
 DR CAROHD: 507 507  
 DR CAROHD: 546 546  
 DR CAROHD: 630 630  
 DR CAROHD: 735 735  
 DR CAROHD: 1116 1116  
 DR CAROHD: 1198 1198  
 DR SEQUENCE: 1331 AA; 14816 MW; CFB2CB55BEFB99C2 CRC64;

Query Match 3.2%; Score 116; DB 1; Length 1331;  
 Best Local Similarity 20.5%; Pred. No. 1.5;  
 Matches 90; Conservative 60; Mismatches 149; Indels 140; Gaps 22;

QY 49 PDGSRMAFV-----SVELCMISQRLLEVLKRLSKYGLSRDNYLSAIHHS 99  
 DB 335 PSSSRKKNKMGESINYNNGVITDLARKKLE-----PSNNG-----NLSFSCVEPYT 373  
 QY 100 GAGGEFOYLYT-----LASGEFSNRT-----FOYVSGIMKSIDIAHTNKP 142  
 DB 374 VPV-FFNATSYLEVGRNLNODLFVSFOFTNPNGLVFSHPADN-IGNVEIDLTESKV 431  
 QY 143 GRIPIKNGVNAVQINRSPSYLLNPOSERARYSSNTDKEMLVLKLVLDNGEDLGLIS-- 200  
 DB 432 G-VHININOTKNSOIDSISGSLNCGOMHEVRFLA---KENAIIITID--GDEASAVRIN 485

QY 201 -----WFAIHPVSMNNSH-----FVNSDNMGIAAYLFQEKKNKGYL 237  
 DB 486 SPLQVTKGKGYFFGFLNMNNSSHSLQPSFGCGMQLQYDDQVNIYEVAQR----- 540  
 QY 238 PGSGPFVAGFASNSNGDVSFNILGPHCVTG---ESCDNDKSTCPNGGSPKMGASGPGD 294  
 DB 541 PGSFANVSIDMCATIDRCVFN---HCEHGKCSOTWDSFKCTCDETGIS-----GAT 589  
 QY 295 MFESTHIIIGRIYQAKELYASASQ-----EVTGPFVLAHQMVNNTDVSVOLNATH 345  
 DB 590 CHNS-----IYPSCEAYKHLGQTSNYWIDPPGSGPLGKYVCNMTEDKVMYIYSH 642  
 QY 346 TVKTCRPAIGS-----FACTIDVSGSLNITQ-----GTBBD 379  
 DB 643 DLQMQTPVGVNPEKYSVTQVLVYSASMDQISAITDSAECEQVYSFCKMSRLMTPPGS 702  
 QY 380 PF--MDTLNDQLLGRKPEE 396  
 DB 703 PTTWM-----VGKANEK 714

RESULT 2  
 CO2\_HUMAN  
 ID CO2\_HUMAN STANDARD: PRT; 752 AA.  
 AC P06681; Q13904;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Complement C2 precursor (EC 3.4.21.43) (C3/C5 convertase).  
 GN C2.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=87127920; PubMed=2949737;  
 RA Bentley D.R.;  
 RT "Primary structure of human complement component C2. Homology to two  
 RT unrelated protein families.";  
 RL Biochem. J. 239:339-345(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93315833; PubMed=8326124;  
 RA Ishii Y., Zhu Z.B., Macon K.J., Volanakis J.E.;  
 RT "Structure of the human C2 gene.";  
 RL J. Immunol. 151:170-174(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,  
 RA Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;  
 RT "Sequence determination of 300 kilobases of the human class III MHC  
 RT locus.";  
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 587-752 FROM N.A.  
 RX MEDLINE=84144868; PubMed=6199794;  
 RA Bentley D.R., Porter R.R.;  
 RT "Isolation of cDNA clones for human complement component C2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:1212-1215(1984).  
 RN [5]  
 RP SEQUENCE OF 694-752 FROM N.A.  
 RX MEDLINE=87102880; PubMed=3643061;  
 RA Wu L., Morley B.J., Campbell R.D.;  
 RT "Cell-specific expression of the human complement protein factor B  
 RT gene: evidence for the role of two distinct 5'-flanking elements.";  
 RL Cell 48:331-342(1987).  
 RN [6]  
 RP SEQUENCE OF 137-171; 454-466 AND 574-717.  
 RX MEDLINE=85038851; PubMed=6149575;  
 RA Gagnon J.;  
 RT "Structure and activation of complement components C2 and factor B.";

RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 306:301-309(1984).  
RN [7]  
RP SEQUENCE OF 244-269.  
RX MEDLINE=83308518; PubMed=6555044;  
RA Parkes C., Gagnon J., Kerr M.A.;  
RT "The reaction of iodine and thiol-blocking reagents with human  
amino acid sequence of a peptide from C2A containing a free thiol  
group.";  
RT Biochem. J. 213:201-209(1983).  
RL [8]  
RN VARIANTS C2D PHE-209 AND ARG-464.  
RX MEDLINE=96215049; PubMed=8621452;  
RA Wetsel R.A., Kulics J., Lokki M.L., Klepela P., Akama H.,  
RA Johnson C.A., Densen P., Colten H.R.;  
RT "Type II human complement C2 deficiency. Allele-specific amino acid  
substitutions (Ser189 --> Phe; Gly444 --> Arg) cause impaired C2  
secretion.";  
RT J. Biol. Chem. 271:5824-5831(1996).  
RN [9]  
RP VARIANT C2D TYR-131.  
RX MEDLINE=98334005; PubMed=9670930;  
RA Zhu Z.B., Atkinson T.P., Volanakis J.E.;  
RT "A novel type II complement C2 deficiency allele in an African-  
American family.";  
RL J. Immunol. 161:578-584(1998).  
CC -1- FUNCTION: COMPONENT C2 WHICH IS PART OF THE CLASSICAL PATHWAY OF  
THE COMPLEMENT SYSTEM IS CLEAVED BY ACTIVATED FACTOR C1 INTO TWO  
FRAGMENTS: C2B AND C2A. C2A, A SERINE PROTEASE, THEN COMBINES WITH  
COMPLEMENT FACTOR 4B TO GENERATE THE C3 OR C5 CONVERTASE.  
CC -1- CATALYTIC ACTIVITY: CLEAVES C3 IN THE ALPHA-CHAIN TO YIELD C3A AND  
C3B. CLEAVES C5 IN THE ALPHA-CHAIN TO YIELD C5A AND C5B. BOTH  
CLEAVAGES TAKE PLACE AT THE C-TERMINAL OF AN ARGININE RESIDUE.  
CC -1- DISEASE: DEFECTS IN C2 ARE THE CAUSE OF C2 DEFICIENCY (C2D). THIS  
IS AN AUTOSOMAL RECESSIVE DISEASE. DEFICIENT INDIVIDUALS HAVE AN  
INCREASED INCIDENCE OF SLE AND SLE-LIKE SYNDROMES.  
CC GLOMERULONEPHRITIS, VASCULITIS AND PYOGENIC INFECTIONS. TYPE I C2D  
IS CHARACTERIZED BY COMPLETE LOSS OF THE PROTEIN WHILE TYPE II C2D  
IS CHARACTERIZED BY A SELECTIVE BLOCK IN C2 SECRETION.  
CC -1- MISCELLANEOUS: C2 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III  
PROTEIN.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -1- SIMILARITY: CONTAINS 3 SUSHI (SCR) DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M15082; AAB59624.1; .  
DR EMBL; L09708; AAB97607.1; .  
DR EMBL; L09706; AAB97607.1; JOINED.  
DR EMBL; L09707; AAB97607.1; JOINED.  
DR EMBL; AF019413; AAB67975.1; .  
DR EMBL; X04481; CAA28169.1; .  
DR PIR; A25290; C2HU.  
DR PIR; A05289; A05289.  
DR HSSP; P00734; 1B7X.  
DR MEROPS; S01.194; .  
DR GeneW; HGNC:1248; C2.  
DR MIM; 217000; .  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR InterPro; IPR002035; VWFA\_A.  
DR Pfam; PF00084; sushi; 3.  
DR Pfam; PF00089; trypsin; 2.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00032; CCP; 3.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS02404; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR PROSITE; PS02344; VWFA; 1.  
KW Complement pathway; Plasma; Glycoprotein; Hydrolase; Serine protease;  
KW Signal; Repeat; Sushi; Disease mutation; Polymorphism.  
FT SIGNAL 1 20  
FT CHAIN 21 752  
FT CHAIN 21 243  
FT CHAIN 244 752  
FT DOMAIN 23 85  
FT DOMAIN 88 145  
FT DOMAIN 150 205  
FT DOMAIN 254 452  
FT DOMAIN 466 752  
FT ACT\_SITE 507 507  
FT ACT\_SITE 561 561  
FT ACT\_SITE 679 679  
FT DISULFID 24 64  
FT DISULFID 51 84  
FT DISULFID 89 131  
FT DISULFID 117 144  
FT DISULFID 151 191  
FT DISULFID 177 204  
FT CARBOHD 29 29  
FT CARBOHD 112 112  
FT CARBOHD 290 290  
FT CARBOHD 333 333  
FT CARBOHD 467 467  
FT CARBOHD 471 471  
FT CARBOHD 621 621  
FT CARBOHD 651 651  
FT VARIANT 131 131  
FT VARIANT 209 209  
FT VARIANT 464 464  
FT VARIANT 533 533  
FT VARIANT 533 533  
SQ SEQUENCE 752 AA; 83267 MW; 5A96A13E700CE444 CRC64;  
Query Match 3.0%; Score 109; DB 1; Length 752;  
Best Local Similarity 21.6%; Pred. No. 2.2;  
Matches 93; Conservative 58; Mismatches 151; Indels 128; Gaps 23;  
Db 133 PGMWDETAVCDGAGCHPFGISL-----GAVTGRFGH-GKVRKCSNLYLVGSS 187  
Qy 318 SEQVYGLVLAHQVNMVDSVQLNATHVTKCPALGYSFACITDVGSLNTGCTE 377  
Db 188 ERECQNGV-----WSGTEPIOROPYSDVDPVAPALGTSS-----HMLGATNPTQKTKE 239  
Qy 378 GDPFWDLRDLQDLKPKSE-----EIVECKPKP-----ILLHSGELTIHPQPD 422  
Db 240 S-----LARKIQLORSGHLNLYLLDCSOSVSENDLIFKESASLMDVRISFE 288  
Qy 423 I-VDOVIVTVGSLAIAIPGELTMSGRPFKAIKKEFALYGMGMTVITAGLSNVYTHY 481  
Db 289 INVSVAIIT-----FASBPVKLVSLVLDNSR-----DMTEVSSLEN----- 325  
Qy 482 ITTYEEYQ-----AQRVEASTY-----GPHLSAVIQLFRDLANAI--ATDVA 525  
Db 326 -ANYKHENGTGTVAALNSVLYMMNMOMQLLGMETMA-----WGEIRHAIILLTQDGS 379  
Qy 526 NMSSGPEPPFFKNIASLIPIDRAPYIGKHFQGVLOPAPAEVY-----EYEV 576  
Db 380 NMGSPPK-----TAVDHIREILINOKRNDYLD-----IYAIGKLDVDMRELNEL 426

QY 577 IFVGNPKNSAENQ-----HQT---LTVKEDSVADMQIMYNDASWETRYWH 624  
DB 427 -----GSKDGERHAFLODTKALHQVFHEMLDVKLTDTICGVNMSANASDOERTPMH 481  
QY 625 KGILGSLNAT 634  
DB 482 VTIKPKSOET 491

RESULT 3  
BIGA\_SALTY STANDARD; PRT; 1953 AA.  
AC P25927; P25928; Q9XC03;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative surface-exposed virulence protein biga precursor.  
OS BIGA OR STM3478.  
GN *Salmonella typhimurium*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX *Salmonella*.  
NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 14028;  
RA Stojiljkovic I., Valentine P., Heffron F.;  
RT "Salmonella typhimurium rbs homolog."  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LT2 / SCS01412 / ATCC 700720;  
RX MEDLINE=21354948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
LT2."  
RL Nature 413:852-856(2001).  
RN [3]  
RP SEQUENCE OF 1-765 FROM N.A.  
RC STRAIN-LT2;  
RX MEDLINE=91100301; PubMed=1987123;  
RA Wu J.Y., Siegel L.M., Kredich N.M.;  
RT "High-level expression of *Escherichia coli* MDPH-sulfite reductase:  
requirement for a cloned cyst plasmid to overcome limiting streheme  
cofactor."  
RT J. Bacteriol. 173:325-333(1991).  
RL -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
in positions 414 and 732.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AF13666; AAD39458.1; -  
DR EMBL: AE008859; AAL22340.1; -  
DR EMBL: M64606; AAA27042.1; ALT\_FRAME.  
DR EMBL: M64606; AAA27043.1; ALT\_FRAME.  
DR PIR: C39200; C39200.  
DR PIR: D39200; D39200.  
DR StyGene: SG10437; biga.  
KW Virulence; Repeat; Signal; Complete proteome.  
FT SIGNAL 27  
FT CHAIN 28 1953  
FT DOMAIN 101 252  
FT REPEAT 101 103 1 (INCOMPLETE).  
FT 15 X 11 AA TANDEM REPEATS.

FT REPEAT 104 113 2 (INCOMPLETE).  
FT REPEAT 114 122 3 (INCOMPLETE).  
FT REPEAT 123 133 4.  
FT REPEAT 134 144 5.  
FT REPEAT 145 155 6.  
FT REPEAT 156 166 7.  
FT REPEAT 167 177 8.  
FT REPEAT 178 188 9.  
FT REPEAT 189 199 10.  
FT REPEAT 200 210 11.  
FT REPEAT 211 221 12.  
FT REPEAT 222 232 13.  
FT REPEAT 233 243 14.  
FT REPEAT 244 252 15 (INCOMPLETE).  
FT CONFLICT 207 207 D -> DRGDDVTPPD (IN REF. 1).  
FT CONFLICT 514 514 A -> R (IN REF. 3).  
FT CONFLICT 1698 1698 D -> N (IN REF. 1).  
FT CONFLICT 1795 1798 OYLE -> ITLO (IN REF. 1).  
FT CONFLICT 1836 1837 SA -> T (IN REF. 1).  
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

Query Match 3.0%; Score 109; DB 1; Length 1953;  
Best Local Similarity 19.0%; Pred. No. 9.8;  
Matches 119; Conservative 86; Mismatches 238; Indels 184; Gaps 26;

QY 3 GYIYGRADCTGOVS---DINMGYKNGQNRGLTFLESPAFILADPDGSGNRMAFVS 59  
DB 859 GYV-SYVRADNGSEVNSGDLIVATYSNSEDRAATRASGSAVHNKAGD-LTLLS 916  
QY 60 VELCMISQRLRLVLR-----RLESKGSLRRDNVLSLHSHSPAGFOYTLVI 111  
DB 917 DQPPGSGGIEVPLWYTFTEYAMASDGDVNDGAT-----HLOAGVYGV--- 968  
QY 112 LASEGFSNRTFOYIVSGIMKSIDIAHTNLKPKIFINKGVANVQINRSPSYLLNPOSE 171  
DB 969 -ASRGKALNCGNITLYDGLVFTLD-----DENNTISTYWPQSLYLS----- 1010  
QY 172 RARYSSNTDKEMLVKLVDNGEDGLISWFAIHPVSMNNSNHFVNSDNKGYAAYLFEOE 231  
DB 1011 -----SGVVASTDADCD-----ATAINGNITVNNAGCGMALNGCYA 1049  
QY 232 KNGYL-----PGGPPVAFGASSN-----LGVSPNIIIG 261  
DB 1050 INGVITLTPADGCVTGADELVGMAALNGCVVINDISGVINIDADYGOAFLDSSSYITIN 1109  
QY 262 PNCVNTGESCDNDKSTCPNGPS--MCMASGPGQ-DMFESTHILGRITLYOKAKELYASAS 318  
DB 1110 NGSINLNGSPMDPTDSHMGCTPTDKTIWQSLPESGDSPTFSDTG---PFTAGTLANYGT 1166  
QY 319 QEYTGPLYLAHQV--NMTDVSVOLNATHYKTKCPALGYSFAGGTIDVSGNLITOGTTE 377  
DB 1167 ETINGDVVNGGWLNEAGASLTVNGT-----VTINGCANALANYGTLD 1210  
QY 378 GDFP--WDTLRDOLLGKPESEIECQKPKRILHSGELTIPHPQDIYDVQIVTGVSLA 435  
DB 1211 ADAISTHSLFENADSGITD-----LTLNDGVTFYN----- 1243  
QY 436 IAAIPBELT-TMNGRRPREAIKKEFALYKGMKDMTVIAGISVYTHYITTYEYQAOQRE 494  
DB 1244 -----NSDFTGSIAGTSYQOEI-----VNTGDMTVLEDKSLVSGSFY-FYNEEDATLTIN 1292  
QY 495 AASTIYGPHLSAYIQFLRAKAIATDVANNSSGP-----EPPFER 537  
DB 1293 SGSAVSGGE--NTIINLTR-----ANDSLTOVSGTITATNGYSAITTVNGSNDPKWIM 1344  
QY 538 NLIASLIPNIADRAPI-----GKHFGD 559  
DB 1345 NTATGVINGINPDAPLNLGRGYNFGN 1371

RESULT 4  
FLID\_BORBU STANDARD; PRT; 665 AA.  
ID FLID\_BORBU



CC - FUNCTION: AGGLUTININATES ERYTHROCYTES.  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY C25.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U41807; AAB17128.1; -  
 DR Interpro: IPR001769; Peptidase\_C25.  
 DR Pfam: PF01364; Peptidase\_C25; 6.  
 DR Hemagglutinin; Virulence; Hydrolase; Thiol protease; Signal; Repeat.  
 KM SIGNAL 1 24  
 FT CHAIN 25 2628 HEMAGGLUTININ A.  
 FT DOMAIN 25 539 PEPTIDASE C25-LIKE 1.  
 FT DOMAIN 540 995 PEPTIDASE C25-LIKE 2.  
 FT DOMAIN 996 1451 PEPTIDASE C25-LIKE 3.  
 FT DOMAIN 1452 1907 PEPTIDASE C25-LIKE 4.  
 FT DOMAIN 2074 2628 PEPTIDASE C25-LIKE 5.  
 SQ SEQUENCE 2628 AA; 283324 MW; 61C4DE32540C99DA CRC64;  
 Query Match 3 0%; Score 106.5; DB 1; Length 2628;  
 Best Local Similarity 18.8%; Pred. No. 24; Mismatches 183; Gaps 31;  
 Matches 116; Conservative 82; Indels 183; Gaps 31;  
 QY 101 PAGF---FOYLYLILASEGFSNRFOYVSGIMK-----SIDIAHTNLKPKIFL--- 147  
 DB 105 PAGVLDPEYKPVANADASFPTNF--VLDGASADIPAGTYIINNPILIIYVGRG 162  
 QY 148 -NKGVANVQINRSSSTILNPQSRARYSSNTDKEMLVKLVDNGEDGLISMPALHP 206  
 DB 163 VSKGNDYVVEAK---TYHFTVQROGP-----GDAASYV----- 193  
 QY 207 VSMNNSNHFVNSDNGYAAVLFEQKKNKGYLGGPFPVAFASNLGDVSPILPHCVN 266  
 DB 194 VTGEGNEFAPVQNLQMSV-----SQGVTLVWQAPASDKRYVL----- 233  
 QY 267 TGESCDNDKSTCPNGSPSCMAAGPGDMFESTHIGRIYOKAKELVASQSEVTGPLY 326  
 DB 224 -NESDFT--QTLPNQM-TMIDAGDGHMLSTIN-----YNTAHTDGC-AM 276  
 QY 327 AAHQVNMNTDVSVQNLNATHYKCPA-----LQSFPAAGTIIDGVS---GLNTOGTE 377  
 DB 277 FSKSWTASGAKIDLSPDNYLVTPTVTPENGKLSY-WVSSQVPTNEHYGVFLSTGNE 335  
 QY 378 GDPEDTLRDQLLGRSEIIVCOKPKPI-LIHSELTPIHMQODIYDVQVIGSLAI 436  
 DB 336 ANNFYIKLEETLGS-----DKPAPNVLKSGVKKLPAPYQERTID-----L 377  
 QY 437 AAIPELTTMSGRRRREAIKKEFALYGMKMTVTVAGLSNVYTHITTYEEOQAOYEVA 496  
 DB 378 SAVAGQOYVLAFRHNSR--GIFRLY-LDDVAVSEGSNDYTY--TYRDMVVAQNLA 432  
 QY 497 STTY-----GPHLSAVIQLFRDLAKAIATDVANNSSGPEPFKNLASLIPNIAD 549  
 DB 433 ATTFQENAVPQOYNYCEVKKYTAGVSPKCVKDYVEGSGN--EFAHVONLITGS----- 483  
 QY 550 RAPICKHFGDVLQPAKPEYRVEGVEYVLFVGANPRNSAENQHOFLIYVEKEDSV-ADM 608  
 DB 484 -----AVGQKVTLMW-----DADGNPNPNPGTTLTISESENGIPIPSW 521  
 QY 609 QIMYND---SWETRFYHKKGLIGLSNATIYWHIDTAVPGIYRIYFGHNKQKOLLAPA 665  
 DB 522 KTIADDGGMNMTT-----TPPPG--GISFAGHN-SALCASSA 556  
 QY 666 VILAFEGISSPEEVTT 682  
 DB 557 SYINFEQGPNPDNYLVT 573

RESULT 6  
 ID CRB1\_HUMAN STANDARD; PRT; 1376 AA.  
 AC P82279;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Crumbs protein homolog 1 precursor.  
 GN CRB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND VARIANTS RP12 V-161, W-250, M-745, C-764;  
 RP Y-948; T-1041 AND P-1071.  
 RC TISSUE=Retina, and Retal brain;  
 RX MEDLINE=9438399; PubMed=10508521;  
 RA van den Born L.I., van Driel M.A., de Kok Y.J.M., van Soest S.,  
 RA Bhattacharya S.S., Kellner U., Hoyng C.B., Westerveld A.,  
 RA Brunner H.G., Bleeker-Wagemakers E.M., Deutman A.F.,  
 RA Heckenlively J.R., Cremers F.P.M., Bergen A.A.B.;  
 RT "Mutations in a human homologue of drosophila crumbs cause retinitis  
 RT pigmentosa (RP12)".  
 RT Nat. Genet. 23:217-221(1999).  
 RL (2)  
 RP VARIANT LCA ARG-1100, AND VARIANT RP ARG-1181.  
 RX MEDLINE=21303018; PubMed=11389483;  
 RA de Kok Y.J.M., van der Velde-Vissers S.D., Kellner U., Jankies B.,  
 RA van Schooneveld M.J., Blankenagel A., Kohnschneider K., Wissing B.,  
 RA Cruysberg J.R.M., Deutman A.F., Brunner H.G., Apfelstedt-Sylla E.,  
 RA Hoyng C.B., Cremers F.P.M.;  
 RT "Leber congenital amaurosis and retinitis pigmentosa with Coats-like  
 RT exudative vasculopathy are associated with mutations in the crumbs  
 RT homologue 1 (CRB1) gene".  
 RL Am. J. Hum. Genet. 69:198-203(2001).  
 CC - FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTION IN NEURONAL  
 CC DEVELOPMENT OF THE RETINA.  
 CC - SUBCELLULAR LOCATION: Extracellular (Potential).  
 CC - TISSUE SPECIFICITY: PREFERENTIAL EXPRESSION IN RETINA, ALSO  
 CC EXPRESSED IN BRAIN AND FETAL BRAIN.  
 CC - DISEASE: DEFECTS IN CRB1 ARE THE CAUSE OF RETINITIS PIGMENTOSA  
 CC TYPE 12 (RP12), AN AUTOSOMAL RECESSIVE CONDITION WHICH IS  
 CC CHARACTERIZED BY NIGHT BLINDNESS FROM EARLY CHILDHOOD AND  
 CC PROGRESSIVE VISUAL FIELD LOSS. THERE IS GENERAL LOSS OF RETINAL  
 CC PIGMENT EPITHELIUM THROUGHOUT THE RETINA AND PATIENTS EXPERIENCE  
 CC SEVERE VISUAL IMPAIRMENT BEFORE THE AGE OF TWENTY.  
 CC - DISEASE: Defects in CRB1 are a cause of a form of Leber congenital  
 CC amaurosis (LCA).  
 CC - SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.  
 CC - SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.  
 CC - DATABASE: NAME-Mutations of the CRB1 gene;  
 CC NONE-Retina international's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci-news/crdmut.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF154671; AAF01361.1; -  
 DR HSSP: P08709; 1BR9.  
 DR Genew; HGNC:2343; CRB1.  
 DR MIM: 604210; -  
 DR MIM: 600105; -  
 DR Interpro: IPR000152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR00742; EGF\_2.





```

OY 287 MASGPDQMFESTHITIGRIYOKAKELYASASOETGPVLAHQWNNMDVSVOLNATPH 346
DB 901 -----VCHSRWDFSCSCPALYTGKACEEYQWCGFSP-----CPHG 936
OY 347 VKTKCPAL-GYSPAGTI-DGYSGL-----NITOGTT 376
DB 937 AQ-COPVLOGFECIANAVFNGSGQILFRSNGNITRELT 974

RESULT 7
FLU1_BACSU
ID FLU1_BACSU STANDARD: PRT: 440 AA.
AC P23445;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Flagellum-specific ATP synthase (EC 3.6.3.14).
GN FLU1
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91258343; PubMed=1828465;
RA Albertini A.M., Carmori T., Crabbs W.D., Scoffone F., Galizzi A.;
RT "The flaA locus of Bacillus subtilis is part of a large operon coding
RT for flagellar structures, motility functions, and an ATPase-like
RT polypeptide."
RT J. Bacteriol. 173:3573-3579(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borstis R., Bourcier L., Brans A., Brun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritsch C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giesepf H., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portolillo S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolet C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Mamuti R., Medler E., Medler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
RL - FUNCTION: PROBABLE CATALYTIC SUBUNIT OF A PROTEIN TRANSLOCASE FOR
CC FLAGELLUM-SPECIFIC EXPORT, OR A PROTON TRANSLOCASE INVOLVED IN
CC LOCAL CIRCUITS AT THE FLAGELLUM.
CC - CATALYTIC ACTIVITY: ATP + H(2O) + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC - SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC - SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC -----
DR EMBL: X56049; CA39523.1; -
DR EMBL: 299112; CAB13497.1; -
DR PIR: S14497; PWBASAS.
DR PIR: D42365; D42365.
DR Subtilisin; BG10243; flil.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR004100; ATPase_a/bn.
DR InterPro: IPR000194; ATPase_a/bcentre.
DR Pfam: PF00006; ATP-synt_ab.1.
DR Pfam: PF02874; ATP-synt_ab_N.1.
DR SMART: SM00382; AAA; 1.
DR TIGRPFAM: TIGR01026; flil_lysN; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR Hydrolase; Hydrogen ion transport; ATP synthesis; ATP-binding;
KW Transport; Protein transport; flagella; Complete proteome.
FT NP_BIND 167 174
FT SEQUENCE 440 AA; 47870 MW; 082F2085AF5D8A4D CRC64;
SQ
Query Match 2.9%; Score 105.5; DB 1; Length 440;
Best Local Similarity 23.0%; Pred. No. 18;
Matches 59; Conservative 40; Mismatches 67; Indels 91; Gaps 15;
OY 357 SEPAQTIDVSG--LNTTGTEGDPFWTLRDQL--GKSEIIVECQKRPILHSQ 411
DB 80 SIAPGSIVETGSLKVKGT-----GLIGQVIDARGEPIDSEF--CRKSPV---ST 127
OY 412 ELTIHPW-QPDL-----VDYQIVTVG----- 432
DB 128 EOSPPNPKRPPIPREMGVQNSID-SLTVGKGRIQIFAGSGVCKSTLMGIAKQTEA 186
OY 433 SLTAIAIPSELTTMSGRFREAIKKEFALYGMKDMTVIAGLSNVTYITTYEYQAO 492
DB 187 DLNVIALVGE---RREREVEFLKDLGKGLKRSIVVA-----TSDQPALMR 231
OY 493 YENASTITGPHLSATIQLEFRLKAI--ATDVANMS-----SGEPPPEFKL--- 539
DB 232 LKNA-----YTATAAEYFRKGQVMPMDSVTFVMAOREIGLAGEPPTTKGTPS 285
OY 540 IASLIPNIDRAPICGH 556
DB 286 VFAILPRLEKRGANEH 302

RESULT 8
XDH_RAT
ID XDH_RAT STANDARD: PRT: 1330 AA.
AC P22965; O63157.
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Xanthine dehydrogenase/oxidase [includes: Xanthine dehydrogenase
DE (EC 1.1.1.204) (XO); Xanthine oxidase (EC 1.1.3.22) (XO) (Xanthine
DE oxidoreductase)].
GN XDH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=90354396; PubMed=2387845;
RA Amaya Y., Yamazaki K.-I., Sato M., Noda K., Nishino T., Nishino T.;
RT "Proteolytic conversion of xanthine dehydrogenase from the

```

RT MAD-dependent type to the O2-dependent type. Amino acid sequence of  
RT rat liver xanthine dehydrogenase and identification of the cleavage  
RT sites of the enzyme protein during irreversible conversion by  
RT trypsin.".  
RL J. Biol. Chem. 265:14170-14175(1990).  
RN (2)  
RP SEQUENCE OF 1-54 FROM N.A.  
RC STRAIN-Sprague-Dawley;  
RA MEDLINE-94268906; PubMed-8208609;  
RX Chow C.W., Clark M., Rinaldo J., Chalkley R.;  
RT "Identification of the rat xanthine dehydrogenase/oxidase promoter.";  
RL Nucleic Acids Res. 22:1846-1854(1994).  
CC -1- FUNCTION: THIS ENZYME CAN BE CONVERTED FROM THE DEHYDROGENASE FORM  
CC (D) TO THE OXIDASE FORM (O) IRREVERSIBLY BY PROTEOLYSIS OR  
CC REVERSIBLY THROUGH THE OXIDATION OF SULFHYDRYL GROUPS.  
CC -1- CATALYTIC ACTIVITY: xanthine + NAD(+) + H(2)O = urate + NADH.  
CC -1- CATALYTIC ACTIVITY: xanthine + H(2)O + O(2) = urate + H(2)O(2).  
CC -1- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Peroxisomal.  
CC -1- INDUCTION: BY INTERFERON.  
CC -1- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
CC -1- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: J05579; AAA42349.1; -  
DR EMBL: U08122; AAA18869.1; -  
DR EMBL: U08120; AAA18869.1; JOINED.  
DR EMBL: U08121; AAA18869.1; JOINED.  
DR PIR: A37810; A37810.  
DR HSSP: P80457; 1FIO.  
DR InterPro: IPR002888; 2Fe-2S bind.  
DR InterPro: IPR000564; 2Fe2S-ferredoxin.  
DR InterPro: IPR000674; Aldxan\_dh\_C.  
DR InterPro: IPR005107; CO\_dexh\_flav\_C.  
DR InterPro: IPR000572; Euk\_Mb\_oxred.  
DR InterPro: IPR001041; Ferredoxin.  
DR InterPro: IPR002346; dehydrog\_molylb.  
DR Pfam: PF00111; fer2\_1.  
DR Pfam: PF00941; FAD\_binding\_5; 1.  
DR Pfam: PF01315; Ald\_xan\_dh\_C; 1.  
DR Pfam: PF01793; fer2\_2; 1.  
DR Pfam: PF02738; Ald\_xan\_dh\_C2; 1.  
DR Pfam: PF03450; CO\_dexh\_flav\_C; 1.  
DR Prodom: PD186071; 2Fe-2S\_bind; 1.  
DR PROSITE: PS00197; 2Fe2S-FERREDOXIN; 1.  
DR PROSITE: PS00559; MOLYBDOPTERIN\_EUK; 1.  
DR Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Iron-sulfur.  
FT INIT\_MET 0  
FT METAL 36  
FT METAL 42  
FT METAL 47  
FT METAL 50  
FT METAL 50  
SQ SEQUENCE 1330 AA; 146111 MW; A3DD206B9D74E565 CRC54;  
Query Match 2.9%; Score 104; DB 1; Length 1330;  
Best Local Similarity 25.6%; Pred. No. 13;  
Matches 79; Conservative 39; Mismatches 107; Indels 84; Gaps 20;  
QY 272 DNDKSTCPNGSPSCMASGP--GQDMFESTHIIIGRIIYQAKKEVYASASQEVTPGLAAH 329  
DB 1064 ETSINTVPMPTAASASADLNGGYEACOTI-----LKRIEPPFKKPTGWEA-- 1114  
QY 330 QWV-NMIDVSYQLAHTVTKCKPALGSPFAGTIIDVSGLNTIYQGTGSDPEWDTLDQ 388  
DB 1115 -WVWDATYSAVLSATGYKT--PNLGYSEF-----TNSGNPF-----HYF 1152

QY 389 ILGKPEEL-VECOCKPRILLHSGELTIPHP-WQPDIVDQIVTWGSLAIAIPGELFTM 446  
DB 1153 SYGACSEVEIDC-----LTDHKNLRTDIY---MDGSSLNRPRI-DIGOV 1194  
QY 447 SGRRFEAIRKKEFALYGMKDMTVIAGLSNVTYHTITTYEEYOARVEASTIYGPHLTS 506  
DB 1195 EG-----AFVQGLGFTMEELHYSPEG--SLHTRGSPYTK-----IPARGISPIE 1237  
QY 507 AYIQLFRLAKAIAFTDIYVANNSSGPEPFEPFNLAISLIPNAD--RAPIGHFGDVLQ-- 562  
DB 1238 FRVSLRLRCPPMKRA--IVASRAVG-EPLLF--LASSIFFAIKDAIRARAQHGNAKOLF 1292  
QY 563 ----PAKPE 567  
DB 1293 QLDSPATPE 1301  
RESULT 9  
PUR6\_DEBOC  
ID PUR6\_DEBOC STANDARD; PRT; 557 AA.  
AC P50504.  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) (AIR  
DE carboxylase) (AIRC).  
GN ADE2.  
OS Debaryomyces occidentalis (Yeast) (Schwanniomyces occidentalis).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
OX NCBI\_TaxID=27300;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-96132031; PubMed-8553700;  
RA Goudron P., Janatova I., Melihoc E., Klein R.D., Costaglioli P.,  
RA Masson J.M.;  
RT "Sequence analysis of the ADE2 gene coding for  
RT phosphoribosylaminoimidazole carboxylase in Schwanniomyces  
RT occidentalis.";  
RL Yeast 11:1289-1293(1995).  
CC -1- CATALYTIC ACTIVITY: 5-amino-1-(5-phospho-D-riboseyl)imidazole-4-  
CC carboxylate = 5-amino-1-(5-phospho-D-riboseyl)imidazole + CO(2).  
CC -1- PATHWAY: De novo purine biosynthesis: sixth step.  
CC -1- SIMILARITY: TO OTHER FUNGAL AND PLANT AIR CARBOXYLASE AND TO THE  
CC AIR CARBOXYLASE PROTEINS, PURE AND PURK, FROM EUBACTERIA.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL: U23210; AAA96380.1; -  
DR HSSP: P09028; 1D7A.  
DR InterPro: IPR000031; AIR\_carboxyl.  
DR InterPro: IPR003135; ATP-grasp.  
DR Pfam: PF00731; AIRC; 1.  
DR Pfam: PF02222; ATP-grasp; 1.  
DR TIGRFAMS: TIGR01161; purK; 1.  
DR TIGRFAMS: TIGR01162; purE; 1.  
DR Purine biosynthesis; Lyase; Decarboxylase.  
KW Purine biosynthesis; Lyase; Decarboxylase.  
SQ SEQUENCE 557 AA; 61338 MW; EE5338090FC2E41A CRC64;  
Query Match 2.9%; Score 103.5; DB 1; Length 557;  
Best Local Similarity 19.2%; Pred. No. 3.7;  
Matches 98; Conservative 92; Mismatches 203; Indels 117; Gaps 25;  
QY 20 INLNGYKNGQANARGLTRFRAFIADP-----DGS-----NRM 55  
DB 6 VGILGGQGLGHMIVYEANRLNIKTVLIDVDPNSPAKQINSNBHVGSGFTDKSIQLAEKC 65

```

QY 56 AFVSVELCMISQRLRLKLEKSYG-SLYRRDNVLSAHTHSGPAGFOYTLYLAS 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 DILVEI-----EHVDYALKYKKEGVEIPLPDITKLQDKYLQKHELIQIGISTYES 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 EGSFNTFOYIVS-----GIMKSIDIAHTLAKPEKTIKNGNVAANOINSPSSYLL 166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 VAVPSNDEQSLEIGKNFNYFMLKSRITLAVD-----GKNVYVVKTESIPEA--L 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 NPOSERARSNN---TDKEMLVKLVDLNGEDLGLISMFALHPVSMNNSHFNV-----S 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 EFLKADPLAEKKCPKPKELAVVVSISEE---VRAIPVETIKHKNICHVYAAPARVS 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 DMNGYAAYLFEQKNGYLPQGQPF-VAGFASSNLDVSPNIGPHCVNTGESCDNDKST 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 DSIAVAKASYLAKNAKSF--PGCGIFGEMFLPN-NELINELIAPRHNSGHTID---- 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 CPNGGSMCMASGPGDMFESTHIIIGRIIYQAKELASASQGYTGVLAHGMVMTDV 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 -----ACVTS-----QFEA-HVRAVVGGLPMKPN-FTSLSTTTNAI-----MLNV 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 338 SVQNLNTHVYKTCRPAKGYSAAGTIDGV-----GLNITQGTGEGDPFMDTLRDOL 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 LGDEENKELEICRRLLETPNASVYLKCKSTRPNRKMGHINIVSSMED---CESRLDYI 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 390 LGKPS---EIVECQKPKPILHSGELTTPHPQPDIVDQIVTVGS--LAIAIDGELT 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 IGKSSKIPENLIPKPKPLVSIIGSDSLP-----VMSVGANILKRGVPELT 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 445 TMSGRFRFAIKKEFALYGMK-DMTVYVING 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 IVSAHRTPHRM--TOYALIEAPKRLKVIING 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
APP2_PLAFA
ID APP2_PLAFA STANDARD: PRT; 451 AA.
AC P13824;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Clustered-asparagine-rich protein (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86206015; PubMed=3517875;
RA Mahlgren M., Aslund L., Franzen L., Sundvall M., Waahlén B.,
RA Berzins K., McNicol L.A., Björntman A., Wigzell H., Perlmann P.,
RA Petersson U.;
RT "A Plasmodium falciparum antigen containing clusters of asparagine
RT residues."
RL Proc. Natl. Acad. Sci. U.S.A. 83:2677-2681(1986).
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13021; AAA29485.1; -
CC PIR: A23535; A23535.
CC InterPro: IPR000504; RMA_rec_mot.
CC Pfam: PF00076; rrm_2.
CC SMART: SM00360; RRM; 2.
CC PROSITE: PS50102; RRM; 2.
CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
CC RNA-binding; Repeat; Malaria.
FT NON_TER 1

```

```

FT DOMAIN 24 114 RNA-BINDING (RRM) 1.
FT DOMAIN 350 443 RNA-BINDING (RRM) 2.
FT DOMAIN 119 128 POLY-ASN.
FT DOMAIN 175 180 POLY-ASN.
FT DOMAIN 195 210 POLY-ASN.
FT DOMAIN 237 245 POLY-ASN.
FT DOMAIN 264 285 POLY-ASN.
SQ SEQUENCE 451 AA; 52204 MW; B686354D85F8C293 CRC64;

Query Match 2.98; Score 103; DB 1; Length 451.
Best Local Similarity 23.48; Pred. NO.2.9;
Matches 51; Conservative 41; Mismatches 74; Indels 52; Gaps 10;

QY 81 YGSLYRRDNVLSA-----ITHSGPAGFOYTL-YILASGFSNRTFOYIVSGIM 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 YLSIYMNNTIVNSTIDVLTNTLHIONIPHTDVLHSLGNGFIK----- 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 KSIDIAHTLAKPEKTIKNGNVAANOINSPSSYLLNPOSERARSNTDKEML-VLK-- 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 52 ---DICYFN--KSKQNNNNNFANKRIYNT-----ALVFETHHEALNVLKNI 94
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 ---LVLDNGEDLGLISMFALHPVSMNNSHFNVDNMGYAAYLFEQ--KNKGYLPQGPF 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 KNILDSGGERNIDAEFAVPNVNINNNNNNSNT-----FQKNMNTNFSQGSTNV 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 VAGFASSNL-GDVSPNIGPHCVNTGESCDNDKSTCPN 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 GSNYSENENFGNNMNNMNYNNSSNNNNNTQNTQN 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
POLS_IBDVA
ID POLS_IBDVA STANDARD: PRT; 1012 AA.
AC P08364;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Structural polyprotein [contains: Major structural protein VP2;
DE Nonstructural protein VP4; Minor structural protein VP3].
OS Avian infectious bursal disease virus (strain Australian 002-73)
OS (IBDV).
OC Viruses; dsRNA viruses; Birnaviridae; Avibirnavirus.
OX NCBI_TaxID=10997;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86259073; PubMed=3014441;
RA Hudson P.J., McKern N.M., Power B.E., Azad A.A.;
RT "Genomic structure of the large RNA segment of infectious bursal
RT disease virus."
RL Nucleic Acids Res. 14:5001-5012(1986).
RN [2]
RP SEQUENCE OF 703-1012 FROM N.A.
RX MEDLINE=86220784; PubMed=3011501;
RA Hudson P.J., McKern N.M., Fahey K.J., Azad A.A.;
RT "Predicted sequence of the host-protective immunogen of infectious
RT bursal disease virus."
RL FEBS Lett. 201:143-146(1986).
CC -1- FUNCTION: SEGMENT A ENCODES A POLYPEPTIDE, THAT IS PROCESSED INTO
CC THE MAJOR STRUCTURAL PROTEINS OF THE VIRION VP2 AND VP3, AND INTO
CC THE PUTATIVE PROTEASE VP4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X03993; CAA27629.1; ALT_INIT.
CC PIR: A24382; GNM3AU.
CC MEROPS: S50.002; -.
CC InterPro: IPR002662; Birna_VP2.

```

DR InterPro: IPR002663; Birna\_VP3.  
 DR InterPro: IPR002664; Birna\_VP4.  
 DR Pfam: PF01766; Birna\_VP2; 1.  
 DR Pfam: PF01767; Birna\_VP3; 1.  
 DR Pfam: PF01768; Birna\_VP4; 1.  
 KW Polypeptide: Structural protein; Nonstructural protein; Hydrolase;  
 FT CHAIN 1 452 MAJOR STRUCTURAL PROTEIN VP2.  
 FT CHAIN 722 NONSTRUCTURAL PROTEIN VP4 (PROTEASE).  
 FT CHAIN 724 1012 MINOR STRUCTURAL PROTEIN VP3.  
 SQ SEQUENCE 1012 AA: 109503 MW: D9320A90459DE86 CRC64;  
 Query Match 2.9%; Score 103; DB 1; Length 1012;  
 Best Local Similarity 20.5%; Pred. No. 10;  
 Matches 104; Conservative 70; Mismatches 185; Indels 148; Gaps 26;  
 QY 231 ENKNGYLPQGGPFAVAFSS-----NLGDVSPNT-LGRHCVTGGSCNDK----- 215  
 DB 153 DRKGNLVGEGVIVLSLPTSYDGLVRLDPIPAIDPKMATCDSSPRRYTTAAD 212  
 QY 276 -----STCPNGGSPMCMASGPGODMEFTHIIRIYOKAKELYASASDEVTPVLAHQ 330  
 DB 213 DVQSSQYOPGCTTILFSA-NIDAITNLVSGELVFQTS-----VQGLVL----- 257  
 QY 331 WNMMDVSVQLANHTVTKCKPALGYSPAAGTIDVSGINTGGTEGDFPMDLRDQL 390  
 DB 258 --NATLYVGFDCT-IVTTRAVAGNGTAGT-DNLMPFLVPTSEIITQPVTSIKLEIV 313  
 QY 391 -----GKPEELVECKRPRIILHSELIT-----PHWOP-DIVIOIVVGSILAT 437  
 DB 314 TSKSGQAGDOM-----SWLASGNLVTHGNYPALRPVTLVNERVATGSAVVV 365  
 QY 438 ALPGELTTSGRFRFAIKKEFALY--GKMDTVI-----AGLSNY-THITTYEEY 468  
 DB 366 AGVSNFELLNPBELANLVTEYGRFPGAMNTKLLSERDRIGITVWPTREYTFPREY 425  
 QY 489 QAQRYEASTIYGPHTLSAYIOLFRLAKAIAITDYANNSGPEPEFKNLIALIPNIA 548  
 DB 426 ----FMEVADLNSPLKIGAFG-FKDIIRIRIAPV-----VASTLEP--- 464  
 QY 549 DRAPRIKHGVDVLPKRP-EYRGEVEVEIF-----VGANPKNSAQNQHOF 595  
 DB 465 -----PAAPLAHALGEGVDYLLGDGAASGTARAKAASGRIRIOL 510  
 QY 596 LVEYKEDSVAD-WQIYNDAWSETRFYMHKGLGSLNATYIHIDVYPGIYRIYRG 654  
 DB 511 LAADKGEVYANLFQVPQNPV-----DGLT-----ASPGVLR--GA 545  
 QY 655 HN-----RKQELKPAVILAFEGISSP 676  
 DB 546 HNLDCVLRGATLFPVITTVEDAMP 572  
 RESULT 12  
 HS60\_SCHPO STANDARD: PRT; 582 AA.  
 ID HS60\_SCHPO 009864; 010285;  
 AC 01-FEB-1996 (Rel. 33, last sequence update)  
 DT 01-FEB-1996 (Rel. 33, last sequence update)  
 DE Heat shock protein 60, mitochondrial precursor (HSP60).  
 GN HSP60 OR MCP60 OR SPAC12612.04.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes;  
 NC NCB1\_TaxID=4896;  
 RN (1)  
 RC SEQUENCE FROM N.A.  
 RX MEDLINE=96144268; PubMed=8566770;  
 RA Yoshida H., Yanaagi H., Yura T.;  
 RT Cloning and characterization of the mitochondrial HSP60-encoding

RT gene of Schizosaccharomyces pombe.";  
 RL Gene 167:163-166(1995).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris P., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,  
 RA Oliver P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA O'Neil S., Pearson D., Quail M.A., Rabbittowisch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skellern J., Simmonds S., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodard J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Welfjans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moesli D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Beyer P., Zimmermann W., Wedler H., Wandut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas R., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Garriti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC - FUNCTION: MAY PARTICIPATE IN ASSEMBLY AND/OR DISASSEMBLY OF  
 CC PROTEINS IMPORTED INTO THE MITOCHONDRION. HSP60 ARE ATPASIS AND  
 CC HAVE AFFINITY FOR UNFOLDED PROTEINS.  
 CC - SUBCELLULAR LOCATION: Mitochondrial.  
 CC - SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
 CC - THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z66568; CA91499.1; -;  
 DR EMBL: D50609; BAA09171.1; -;  
 DR HSSP: P06139; IGRU.  
 DR InterPro: IPR001844; Chaprin Cpn60.  
 DR InterPro: IPR002423; Cpn60/TCP-1.  
 DR Pfam: PF00118; cpn60\_TCP1.1.  
 DR PRINTS: PR00298; CHAPERONIN60.  
 DR PROSITE: PS00296; CHAPERONIN60.  
 KW Chapterone: ATP-binding; Mitochondrion; Transit peptide; Heat shock.  
 FT TRANSIT 1 35 MITOCHONDRION (POTENTIAL).  
 FT CHAIN 36 582 HEAT SHOCK PROTEIN 60.  
 FT CONFLICT 392 392 Y -> S (IN REF. 1).  
 FT CONFLICT 415 415 G -> A (IN REF. 1).  
 SQ SEQUENCE 582 AA: 62167 MW: E5434436A9BAFC05 CRC64;  
 Query Match 2.9%; Score 102.5; DB 1; Length 582;  
 Best Local Similarity 23.4%; Pred. No. 4.7;  
 Matches 62; Conservative 42; Mismatches 94; Indels 67; Gaps 14;  
 QY 133 IDIAHTNLRPKGKIFINKGVANYOINRSPSYL-----LNPOSERAR-----YSSN 178  
 DB 336 IDVSIKKAQPHNL-----GSGSGVTVTKEDPIIIMKAGADHKVNDRCQGVADPNLRE 391  
 QY 179 TDKEMVLKLVLDNGEDLGISFATHPMSM-NSNHPNDSMDMGTAATLFEDEKKKGL 237  
 DB 392 YEKEKLOERLAKLSG-GIAVIRKGSSEVEVNEKKRIVDALNAVAAY-----SEGLV 444

QY 238 PGCG-PPVAGFASNLGDSVPN-----ILGPHCVNTGESCONDKSTCPNGSPSCMASGPG 292  
DB 445 PGAGTSTFVK--ASTLRIGDPTNNFDOKLGEIVRKA-----LTRPA 483  
QY 293 QDMEESTHIIIGRIYOKAKELVYASASQEVTPVLAHOMVMTDVSVOLNATHYKCKRP 352  
DB 484 QTLLENAGLGNILVGLKELCKERN--IGYDIADKDFVLDNIGV-LDPLKVVRT--- 537  
QY 353 ALGVSFAGTIDVSGNLITQGTTE 377  
DB 538 -----GLVD-ASGVASLMTTE 553  
RESULT 13  
STAU\_DROME  
ID STAU\_DROME STANDARD: PRT; 1026 AA.  
AC P25159;  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Maternal effect protein staufen.  
GN STAU.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91300552; PubMed=1712672;  
RA St Johnston D., Beuchie D., Nusslein-Volhard C.;  
RT "Staufen, a gene required to localize maternal RNAs in the Drosophila  
RT egg";  
RL Cell 66:51-63(1991).  
RN [2]  
RP STRUCTURE BY NMR OF 579-646.  
RX MEDLINE=95354674; PubMed=7628456;  
RA Bricot M., Grunert S., Muzlin A.G., Proctor M., St Johnston D.;  
RT "NMR solution structure of a dsRNA binding domain from Drosophila  
RT stau protein reveals homology to the N-terminal domain of  
RT ribosomal protein S5";  
RL EMBO J. 14:3563-3571(1995).  
RN [3]  
RP ERRATUM.  
RA Bricot M., Grunert S., Muzlin A.G., Proctor M., St Johnston D.;  
RL EMBO J. 14:4385-4385(1995).  
RN [4]  
RP CHARACTERIZATION OF DRBM DOMAINS.  
RX MEDLINE=20183617; PubMed=10716936;  
RA Micklem D.R., Adams J., Grunert S., St Johnston D.;  
RT "Distinct roles of two conserved Staufen domains in oskar mRNA  
RT localization and translation";  
RL EMBO J. 19:1366-1377(2000).  
CC -1- FUNCTION: REQUIRED BOTH FOR THE LOCALIZATION OF MATERNAL  
CC DETERMINANTS TO THE POSTERIOR POLE OF THE DROSOPHILA EGG AND  
CC FOR BICOID RNA TO LOCALIZE CORRECTLY TO THE ANTERIOR POLE.  
CC OSKAR PROTEIN IS REQUIRED TO KEEP OSKAR RNA AND STAUFEN PROTEIN AT  
CC THE POSTERIOR POLE.  
CC -1- TISSUE SPECIFICITY: POLAR GRANULES AT THE POSTERIOR POLE OF THE  
CC OOCYTE, AND BY THE TIME THE EGG IS LAID, AT THE ANTERIOR POLE.  
CC -1- DOMAIN: CONTAINS A PROLINE-RICH DOMAIN. THE INSERTION OF THIS  
CC DOMAIN IN THE DRBM 2 DOMAIN IS REQUIRED FOR STAU-OSKAR MRNA  
CC LOCALIZATION.  
CC -1- SIMILARITY: CONTAINS 5 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M69111; AAA73062.1; -  
DR PIR; A40315; A40315.  
DR PDB; 1STU; 31-JUL-95.  
DR FlyBase; FBgn0003520; stau.  
DR InterPro; IPR001159; DS\_RBD.  
DR Pfam; PF00035; dsrm; 5.  
DR SMART; SM00358; DSRM; 4.  
DR PROSITE; PSS0137; DS\_RBD; 5.  
KW Developmental protein; RNA-binding; Repeat; 3D-structure.  
FT DOMAIN 311 378 DRBM 1.  
FT DOMAIN 398 557 DRBM 2 (ATYPICAL).  
FT DOMAIN 578 645 DRBM 3.  
FT DOMAIN 711 781 DRBM 4.  
FT DOMAIN 951 1018 DRBM 5.  
FT DOMAIN 3 77 GLN/HIS/PRO-RICH.  
FT DOMAIN 42 46 POLY-GLN.  
FT DOMAIN 47 51 POLY-PRO.  
FT DOMAIN 71 77 POLY-GLN.  
FT DOMAIN 451 455 POLY-PRO.  
SQ SEQUENCE 1026 AA; 110347 MW; AE5B97624BBF7D0B CRC64;

Query Match 2.9%; Score 102.5; DB 1; Length 1026;

Best Local Similarity 18.4%; Pred. No. 11; Matches 135; Conservative 100; Mismatches 255; Indels 245; Gaps 36;

QY 5 YIGVGRADCTGOVSDINLMGYGRKNGONARGLTLRFSRADIADPGSNMAFVSELCM 64  
DB 283 HVTGNDATGALSNEDTSSSGRGKD-----KTPMCL 315  
QY 65 ISQRLREVLKLESKGSLYRRDNVLSAIHSGPAGFOYTYIL-----ASGFS 118  
DB 316 VNEIARY-----NKTTHQYR-----LTERGPAHCKTQTVTLMDGEYSADGFK 360  
QY 119 NRTFOYIVSGIMKSIDIAHTNLKPGKIFINKGVANVQINRSPSYLLNPOSERARYSN 178  
DB 361 IKKAQHIAAS--KAIEETWKHKPPK-----IKRSEEG--GPM--RTHTTPT 401  
QY 179 TDKEMLVKLVLDNGEDLGLISWFAIHVPMNSN-----HFVNSDNMGY----- 223  
DB 402 VELNALAMK-----LGORTFYLDDPTQIPPTDSIVPPEFAGHILTPAGCPMPQPP 453  
QY 224 -AAYLVEQENKNGYL-----GCGPFFVAGFASNLGDSVPNIIIGPHCVNTG 268  
DB 454 PPAVALRQRLGNGVPIPSQPMHHPFHGQRPFPKPSR--FALPPLLAH-VHGG 509  
QY 269 ESCDNDKSTCPNGSPSCMASGP-----GDMFESTHIIIGRIYOKAKELYA----- 315  
DB 510 -----PN-GPPSVPTPPSKITLFYGRKQKFGV--IGRTIQAKKDAARALQVL 555  
QY 316 -----SASQEV-----TGPVLAHOMVMTDVSVOLNATHYKCK---PALGYS 357  
DB 556 KTOAISASEEALIEDSMDEGDKKSPISQVH-----EIGIKRNMTVHFKVIREGPAHMK 609  
QY 358 FAAGTIDVSGNLITQGTTEGDPFWDLRQOLLGKPSSE--IYECCKPKILLHSGTLT 415  
DB 610 FITACT--VGSITVEGNG-----KVSCKRAAEMLVLEQLPLPLPTPK----- 653  
QY 416 PHPWQPDIVQIVTVGSIAIAIPGL-----TMSGRRFEPAIKKEFALYGMDMTV 469  
DB 654 -----QPLKRIKTKTPKSAARREGSVSGTGTGTGKRPERRK-----RLNPPDKLI 704  
QY 470 VIAGLSNVYTHY--TTEEYQ-----AQRYE-----AASTYGPHT-- 504  
DB 705 DMDADNPITKLQLQOQTRKEKEPIELIANKGNETARRREFMEVSASGSTARCGNSK 764  
QY 505 -----LSAYIOLFDDKAIKAITDVYANNS-----SGPEPPFPKNLIASLIPIADR 550  
DB 765 KLAKRNAQAQLFELLEAVOVTPTNETOSSSECTSATMSAVTAPAVEATAREGVPAVNA-- 822  
QY 551 APIGKHFQDLV-----QPAKPEYRAGVEVIVFVYGANPKASANOHTQTLVTEKEDSV 605  
DB 823 TPVGPMPGILLILNKKPKAKR-----DQIVIVKSVESKEBEAKKEVAVAAEENSNS 876

QY 606 ADMQIMVDSWETR 620  
DB 877 ANSGDSSNSSGSDSQ 891

RESULT 14  
AF10\_HUMAN STANDARD: PRT; 1027 AA.  
ID AF10\_HUMAN  
AC P55197;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE AF-10 protein.  
GN MLLT10 OR AF10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95195207; PubMed=788665;  
RA Chaplin T., Ayton P., Bernard O.A., Saha V., della Valle V.,  
RA Hillion J., Gregorini A., Lillington D., Berger R., Young B.D.;  
RT "A novel class of zinc finger/leucine zipper genes identified from  
the molecular cloning of the t(10;11) translocation in acute  
leukemia.";  
RL Blood 85:1435-1441(1995).  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- TISSUE SPECIFICITY: ABUNDANTLY IN TESTIS.  
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL  
TRANSLOCATION T(10;11)(P12;Q23) THAT INVOLVES MLLT10 AND MLL/HRX.  
CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.  
CC -1- SIMILARITY: HIGH TO AF17.  
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.  
CC -1- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;  
WWW="http://www.intoblogen.fr/services/chromancer/Genes/AF10.html".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; U13948; AAA79972.1; -  
DR Genew; HGNC:16063; MLLT10.  
DR MIM; 602409; -  
DR InterPro; IPR001965; Znf\_PHD.  
DR Pfam; PF00628; PHD; 1.  
DR SMART; SM00249; PHD; 2.  
DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
DR PROSITE; PS50016; ZF\_PHD\_2; 2.  
KW Nuclear protein; Zinc finger; Repeat; Chromosomal translocation;  
KW Proto-oncogene.  
FT ZN\_FING 22 74 PHD-TYPE 1.  
FT ZN\_FING 68 85 C4-TYPE.  
FT ZN\_FING 134 197 PHD-TYPE 2.  
FT DOMAIN 229 240 GLU/LYS-RICH.  
FT DOMAIN 766 794 LEUCINE-ZIPPER.  
FT DOMAIN 856 891 POLY-SER.  
FT SITE 266 266 MLL FUSION POINT (IN ACUTE MYELOID  
LEUKEMIA PATIENT B).  
FT SITE 643 643 MLL FUSION POINT (IN ACUTE MYELOID  
LEUKEMIA PATIENT C).  
FT SITE 680 680 MLL FUSION POINT (IN ACUTE MYELOID  
LEUKEMIA PATIENT A).  
SQ SEQUENCE 1027 AA; 109026 MW; 7C7C0059DD46589 CRC64;  
Query Match 2.9%; Score 102.5; DB 1; Length 1027;  
Best Local Similarity 19.3%; Pred. No. 11;  
Matches 99; Conservative 63; Mismatches 189; Indels 163; Gaps 20;

QY 27 KNGQNAKGLTRLEFRALADPDGSSNMFAVSELMISQRLREVLKRLKLE-----S 79  
DB 485 KNGENVSLSVSSASPTSSVSAAGS-----ITSSSLQSPILLRNGSLQSLVSGSPVGS 540  
QY 80 KYGSLYRQDNV-----ILSAITHGCPAGFFQYLYILASEGFSNRTQYIVSGIMK 131  
DB 541 EISMQYRRHGACPTTFSEELINALHNDGSD-----SSTLTQELKFI--GIYN 586  
QY 132 SIDINHTMLKPKGKIFINKGNVANNVQINRSPSYLLNPOSERARYS-----SNPK 181  
DB 587 SMDVA-----VSFPNVVSGSGSTPVSSSHLPQOSSGHLOQVALSPSAVSSAP 636  
QY 182 EMLVLEKLVNAGEDLGLSWFAIHPEVSNMNSNHEFVSDNMGYAAYLPEQENKGYLPGQG 241  
DB 637 AVATIQAMTLGSSSL-----SQAPSIMYCNRRNSSMAALIAQSENNQ----- 678  
QY 242 PIVAGFASNNLGDVSPNLLPGHCVNTGSCDNKSTCPNG--GPSMCAS-----GPGQ 293  
DB 679 -----TDQDLGDNRNILVG-----RGSSPRGSLSPRSPVSLQIRYDQGN 719  
QY 294 DMFESTHIGRIYQKAKELVASAQ-----EYGPVIAAHQW-----VN 333  
DB 720 SLENLPPVASTLEQLERQNSEGOQLLEQGPSDLIGMLKSLHQLOVENRRLKEQIKN 779  
QY 334 MTD-----VSVOLNATHVTCKPKALG---YSFAAGTIDVSGNLITGTEGDPFMD 383  
DB 780 LTAKEERQLNLAQSVFPPTITANPSPSHQHTHTSATAPTTDBLSNKSIGHNSF-- 837  
QY 384 TLRDQLLKPSSEIVECKPKPILLHSGELTIPH-----WQPDIVQIVTVGS 433  
DB 838 -LPDNLSPVLMQDLTSSGQSTS---SSSALSTPPAGQSPAQGGSGVGVQVGVTVGA 893  
QY 434 LA-----IAIRGELTMSG 448  
DB 894 LASGMQPYTSTIPAVSAVGIIIGLPGQALING 927

RESULT 15  
MTE\_CAMJ2  
ID MTE\_CAMJ2 STANDARD: PRT; 754 AA.  
AC Q9PN94;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 5-methyltetrahydropteroyltryptanate--homocysteine methyltransferase  
(EC 2.1.1.14) (Methionine synthase, vitamin B12 independent isozyme)  
DE (Cobalamin-independent methionine synthase).  
GN MTE OR C1201.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feilwell T., Holtroyd S.,  
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrrell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-  
METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE  
FORMATION (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltry-L-glutamate + L-  
homocysteine = tetrahydropteroyltry-L-glutamate + L-methionine.  
CC -1- COFACTOR: ZINC. BINDS ONE ION PER SUBUNIT (BY SIMILARITY).  
CC -1- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.  
CC -1- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE  
SYNTHASE FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL139077; CAB73455.1; -
DR InterPro: IPR002629; Methionine_synth.
DR Pfam: PF01717; Methionine_synth.1
KW Transferase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW Complete proteome.
FT METAL 641 641 ZINC (BY SIMILARITY).
FT METAL 643 643 ZINC (BY SIMILARITY).
FT METAL 726 726 ZINC (BY SIMILARITY).
SQ SEQUENCE 754 AA; 86922 MW; 113AFB568EFECOD CRC64;

Query Match 2.8%; Score 101.5; DB 1; Length 754;
Best Local Similarity 19.7%; Pred. No. 8.4;
Matches 116; Conservative 84; Mismatches 203; Indels 187; Gaps 31;

OY 174 RYSNNTDKEMLVKLVLDNGEDLGISWFAIHVSMN--NSNHFVNSDMNGYAALFE-- 229
DB 23 KTFKQSSKEELK---SAKDLRIHWQEIQKAGIDFIPSNDFSLYDNLDAVLENTIV 78
OY 230 --QEKV-----KGYLPGGPFVAGFASNLGDVSPNIIQPHCVMTGESCDND 274
DB 79 HRTKYNLNLDALDEYFAQSRGY-QGNGCVTALAMKKPNTNHYLVPECDNAD----- 131
OY 275 KSTCPNGSPGCMASGPGQDMFESTHIIIGRIIYQAKELYASASQEVTPVLAHQVMN 334
DB 132 -----IALTGDKIFKE-----YLEAKELGIESKPVLIQ-IFLFLKLI AF 170
OY 335 TDVSVQNLNTHVTKCKPALGYSFAAGTIDVSGNLITQGTGDPF--WDTLRDQ--LL 390
DB 171 KDEKTO-----KLAKERLNNAYLE-LPDKLNELKVTWLELD-EPYLVYDLSKEDIALF 221
OY 391 GKPSIEIVECQKPKPILHS--GELTIIPHPQPDIVQIVTWSLAIAIPGELTMSG 448
DB 222 EEFYQELLNHKDKILLOSIFGDL--RDIYPKLIESKFDALG-----LDFIEG 268
OY 449 RRFREAIKKEFALYGMKDMTVVIAGISNVTYHTYEE-----YQQRV-----EAA 496
DB 269 KOSLALWQO---YGFAPKDKILFGGLNGKNIYANDYAKSLKLIKELQKTONIILNTSC 324
OY 497 STIYGPHTL-----SAVIOLE-----ROLAKAIATDVANMSSGPEPPFKNL 539
DB 325 SLHVPYSTEFESKLDSSYLKLPAPAKERLOELKDL-KEILNSSEENPLFRANQELFKNI 383
OY 540 IASL-----IPNIADRAPIGKHGVDVLOPAKPEYRVGEVEVIFVGA 581
DB 384 PERLDEKVARLALKLKEDFTKPSFKERALLQKER--LKLPLPTTTIGSFQSDVRS 441
OY 582 N-----PKNSAENQTHQTLVEKEYEDSVADMQIMYNDASWETRFYWHKIGLISNAT 635
DB 442 NRLAFKQEKISAQNYTE--FNQOKIKKECI--QIOEE-----IGL----- 476
OY 636 YVHIDPTAVPGIYR-----IRYFGHNRKQEL-----LKPAYI 667
DB 477 -----DVLVHGFEFRNDVYFEGENLKGLFTQNGWVQSYGTRCVKRPVI 521

```

Search completed: July 3, 2003, 12:43:42  
 Job time : 28 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2003, 12:42:40 ; Search time 90 Seconds  
(without alignments)  
1561.379 Million cell updates/sec

Title: US-09-937-521-14

Perfect score: 3582

Sequence: 1 FSGYIGVGRADCTGVSDI.....KPAVILAFEGISSPEFVVT 682

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: SP\_archaea:\*  
3: SP\_bacteria:\*  
4: SP\_fungi:\*  
5: SP\_human:\*  
6: SP\_invertebrate:\*  
7: SP\_mammal:\*  
8: SP\_mhc:\*  
9: SP\_organelle:\*  
10: SP\_phage:\*  
11: SP\_plant:\*  
12: SP\_rodent:\*  
13: SP\_virus:\*  
14: SP\_vertebrate:\*  
15: SP\_unclassified:\*  
16: SP\_rvtrus:\*  
17: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3582	100.0	756	11	Q9JHE3 mus musculus
2	3416	95.4	761	11	Q91XT9 rat
3	3044	85.0	761	4	Q9NR71
4	2634	73.5	496	11	Q8R236
5	1509.5	42.1	704	5	Q9VAX0
6	1416	39.5	818	10	Q9AX93
7	1398	39.0	808	10	Q9LNV7
8	1373	38.3	702	5	Q15913
9	1249.5	34.9	715	10	Q9SH86
10	1223	34.1	705	10	Q9FTL4
11	1120.5	31.3	841	3	Q8X0X9
12	1095.5	30.6	670	16	Q91596
13	1090.5	30.4	670	2	Q9RH00
14	884.5	24.7	637	16	Q06769
15	626	17.5	314	10	Q93Z16
16	152	4.2	1794	16	Q8RHH1

17	136	3.8	2742	16	Q8XX21	Q8XX21 raltosia s
18	131.5	3.7	1565	2	Q85156	Q85156 photorhabd
19	131.5	3.7	2013	16	Q92EK2	Q92EK2 listeria in
20	131	3.7	616	16	Q8ZEK3	Q8ZEK3 yersinia pe
21	131	3.7	625	16	Q8R976	Q8R976 thermoaer
22	118.5	3.3	2013	16	Q8Y978	Q8Y978 listeria mo
23	116.5	3.3	1417	16	Q9CK92	Q9CK92 pasteurilla
24	115	3.2	732	10	Q9FIG2	Q9FIG2 arabidopsis
25	113.5	3.2	1175	2	Q47955	Q47955 haemophilus
26	113	3.2	2015	16	Q9URD2	Q9URD2 neisseria m
27	112	3.1	3972	16	P73139	P73139 synechocyst
28	111.5	3.1	847	6	Q8S074	Q8S074 pan troglod
29	111.5	3.1	752	10	Q8UQ73	Q8UQ73 human immu
30	110.5	3.1	1167	11	Q88341	Q88341 rattus norv
31	110	3.1	1217	11	P97685	P97685 rattus norv
32	110	3.1	1340	3	Q8TFK0	Q8TFK0 debrayomyce
33	110	3.1	2566	17	Q8TSE7	Q8TSE7 methanosarc
34	109	3.0	577	7	Q951G1	Q951G1 homo sapien
35	109	3.0	1736	10	Q81899	Q81899 arabidopsis
36	109	3.0	1975	16	Q9K057	Q9K057 neisseria m
37	109	3.0	1995	16	Q9JY23	Q9JY23 neisseria m
38	108	3.0	752	6	Q8S075	Q8S075 pongo pygma
39	108	3.0	854	15	Q74459	Q74459 human immu
40	108	3.0	1053	2	Q30776	Q30776 nostoc punc
41	108	3.0	1576	13	Q91908	Q91908 xenopus lae
42	108	3.0	2271	13	Q91909	Q91909 xenopus lae
43	107.5	3.0	677	10	Q95Z56	Q95Z56 arabidopsis
44	107.5	3.0	1193	16	Q97M47	Q97M47 clostridium
45	107.5	3.0	2327	12	Q9QCE4	Q9QCE4 foot-and-mo

## ALIGNMENTS

### RESULT 1

ID	Q9JHE3	PRELIMINARY:	PRT:	756 AA.
AC	Q9JHE3			
DT	01-OCT-2000 (TREMBLrel. 15	Created		
DT	01-OCT-2000 (TREMBLrel. 15,	Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17,	Last annotation update)		
DE	Neutral ceramidase (Neutral	ceramidase)		
GN	ASAH2 OR BCDASE OR LCDASE.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata;	Cranialata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia;	Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RX	MEDLINE=20219171; PubMed=10753931;			
RA	Tani M., Okino N., Mori K., Tanigawa T., Izu H., Ito M.;			
RT	"Molecular cloning of the full-length cDNA encoding mouse neutral			
RT	ceramidase -Novel but highly-conserved gene/family of neutral/alkaline			
RT	ceramidases."			
RL	J. Biol. Chem. 275:11229-11234(2000).			
DR	EMBL; AB037181; BAA94546.1;			
DR	EMBL; AB037111; BAA94545.1;			
DR	MGD; MGI:1859310; Asah2.			
SQ	SEQUENCE 756 AA; 83508 MW; FPD514E51280D4BE CRC64;			

QY	1 FSGYIGVGRADCTGVSDIMMGYKNGQVARGILTRLFSSRAFLADPDGSSNMAFVSV 60	Best Local Similarity 100.0%; Score 3582; DB 11; Length 756;
DB	75 FSGYIGVGRADCTGVSDIMMGYKNGQVARGILTRLFSSRAFLADPDGSSNMAFVSV 134	Matches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	61 ELCAISQRLKLEVLKRLSKGSLYRDNVLLSIRHSGPAGFFQITLYLASEGFSNR 120	
DB	135 ELCAISQRLKLEVLKRLSKGSLYRDNVLLSIRHSGPAGFFQITLYLASEGFSNR 194	

```
QY 121 TFOYIVSGIMKSTDIAHTNLKPKKIFINGGNVANOINSPSSYLLNPOSERRARYSSND 180
DB 195 TFOYIVSGIMKSTDIAHTNLKPKKIFINGGNVANOINSPSSYLLNPOSERRARYSSND 254
QY 181 KEMLVKLVLDNGEDGLISMFAIHPVSMNNSNHFVNSDMGAAVLFDEKKKGLPGQ 240
DB 255 KEMLVKLVLDNGEDGLISMFAIHPVSMNNSNHFVNSDMGAAVLFDEKKKGLPGQ 314
QY 241 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNDKSTCPNGSPGSMASGPGDMESTH 300
DB 315 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNDKSTCPNGSPGSMASGPGDMESTH 374
QY 301 IIGRIYQAKELIASASQEVTPVLAHQVNMVTVSVQLNATHVTKCPKALGYSFAA 360
DB 375 IIGRIYQAKELIASASQEVTPVLAHQVNMVTVSVQLNATHVTKCPKALGYSFAA 434
QY 361 GTIDVSGNLITOGTEGPFMDTLRDQLLGRKSEIEVCOKRPILHSGELTIPHPQ 420
DB 435 GTIDVSGNLITOGTEGPFMDTLRDQLLGRKSEIEVCOKRPILHSGELTIPHPQ 494
QY 421 PDIVDQIVTVGSLAIAIPGELITVSGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 480
DB 495 PDIVDQIVTVGSLAIAIPGELITVSGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 554
QY 481 YTTTVEEYQAEASTIYGPHTLSAYIQLFRDLAKAATDPAVNMSSGPEPFKNLI 540
DB 555 YTTTVEEYQAEASTIYGPHTLSAYIQLFRDLAKAATDPAVNMSSGPEPFKNLI 614
QY 541 ASLIPNIADRAPIGKHFQDVLQPAKPEYRGEVEYIFGANKPNSAENQHOTFLTVK 600
DB 615 ASLIPNIADRAPIGKHFQDVLQPAKPEYRGEVEYIFGANKPNSAENQHOTFLTVK 674
QY 601 YEDSVADMOIMYNDASMETRFYWHKGIILGSNATIWHIPDTAIPGIYIRYFGHNRKOE 660
DB 675 YEDSVADMOIMYNDASMETRFYWHKGIILGSNATIWHIPDTAIPGIYIRYFGHNRKOE 734
QY 661 LKPAVILAFEGISSPEFVYTT 682
DB 735 LKPAVILAFEGISSPEFVYTT 756

RESULT 2
Q91XT9 PRELIMINARY: PRT: 761 AA.
AC 091XT9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE Ceramidase.
GN PAPKCD.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID-10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RX MEDLINE-21336561; PubMed-11328816;
RA Mitsutake S., Tanai M., Okino N., Mori K., Ichinose S., Omori A.,
RA Iida H., Nakamura T., Ito M.;
RT "Purification, Characterization, Molecular Cloning and Subcellular
RT Distribution of Neutral Ceramidase of Rat Kidney.";
RL J. Biol. Chem. 276:26249-26259(2001);
DR EMBL; AB057433; BAB62033.1;
SQ SEQUENCE 761 AA; 83487 MW; 68B91BC78AEB6324 CRC64;
```

Query Match 95.4%; Score 3416; DB 11; Length 761;  
Best Local Similarity 94.1%; Pred. No. 1.5e-262;  
Matches 642; Conservative 25; Mismatches 15; Indels 0; Gaps 0;

QY 1 FSGYIVGVADCTGVSDINLMGYKNGKNGARGLTLRFSAFIADPDGSRMAFVS 60  
DB 80 FSGYIVGVADCTGVSDINLMGYKNGKNGARGLTLRFSAFIADPDGSRMAFVS 139

```
QY 61 ELICMISORLLEVLKRLKESKYSGLYRDDNYILSAIHTSGPAGFOYTYLLASEGFSNR 120
DB 140 ELICMISORLLEVLKRLKESKYSGLYRDDNYILSAIHTSGPAGFOYTYLLASEGFSNR 199
QY 121 TFOYIVSGIMKSTDIAHTNLKPKKIFINGGNVANOINSPSSYLLNPOSERRARYSSND 180
DB 200 TFOYIVSGIMKSTDIAHTNLKPKKIFINGGNVANOINSPSSYLLNPOSERRARYSSND 259
QY 181 KEMLVKLVLDNGEDGLISMFAIHPVSMNNSNHFVNSDMGAAVLFDEKKKGLPGQ 240
DB 260 KEMLVKLVLDNGEDGLISMFAIHPVSMNNSNHFVNSDMGAAVLFDEKKKGLPGQ 319
QY 241 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNDKSTCPNGSPGSMASGPGDMESTH 300
DB 320 GPFVAGFASNLGDVSPNLLGPHCVNTGSCDNDKSTCPNGSPGSMASGPGDMESTH 379
QY 301 IIGRIYQAKELIASASQEVTPVLAHQVNMVTVSVQLNATHVTKCPKALGYSFAA 360
DB 380 IIGRIYQAKELIASASQEVTPVLAHQVNMVTVSVQLNATHVTKCPKALGYSFAA 439
QY 361 GTIDVSGNLITOGTEGPFMDTLRDQLLGRKSEIEVCOKRPILHSGELTIPHPQ 420
DB 440 GTIDVSGNLITOGTEGPFMDTLRDQLLGRKSEIEVCOKRPILHSGELTIPHPQ 499
QY 421 PDIVDQIVTVGSLAIAIPGELITVSGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 480
DB 500 PDIVDQIVTVGSLAIAIPGELITVSGRRFRFAIKKEPALYGMKDMTVIAGLSNVYTH 559
QY 481 YTTTVEEYQAEASTIYGPHTLSAYIQLFRDLAKAATDPAVNMSSGPEPFKNLI 540
DB 560 YTTTVEEYQAEASTIYGPHTLSAYIQLFRDLAKAATDPAVNMSSGPEPFKNLI 619
QY 541 ASLIPNIADRAPIGKHFQDVLQPAKPEYRGEVEYIFGANKPNSAENQHOTFLTVK 600
DB 620 ASLIPNIADRAPIGKHFQDVLQPAKPEYRGEVEYIFGANKPNSAENQHOTFLTVK 679
QY 601 YEDSVADMOIMYNDASMETRFYWHKGIILGSNATIWHIPDTAIPGIYIRYFGHNRKOE 660
DB 680 YEDSVADMOIMYNDASMETRFYWHKGIILGSNATIWHIPDTAIPGIYIRYFGHNRKOE 739
QY 661 LKPAVILAFEGISSPEFVYTT 682
DB 740 LKPAVILAFEGISSPEFVYTT 761

RESULT 3
Q9NR71 PRELIMINARY: PRT: 761 AA.
AC 09NR71;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE Mitochondrial ceramidase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID-9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20347271; PubMed-10781606;
RA El Bawab S., Roddy P., Qian T., Bielawska A., Lemasters J.J.,
RA Hannun Y.A.;
RT "Molecular cloning and characterization of a human mitochondrial
RT ceramidase.";
RL J. Biol. Chem. 275:21508-21513(2000);
DR EMBL; AF250847; AAF86240.1;
SQ SEQUENCE 761 AA; 83239 MW; 04F881ED34A27E4 CRC64;
```

Query Match 85.0%; Score 3044; DB 4; Length 761;  
Best Local Similarity 82.2%; Pred. No. 5.7e-233;  
Matches 560; Conservative 63; Mismatches 58; Indels 0; Gaps 0;

Qy	1	FSGYIGVGRADCTGOVSDINIMMGYKGNQNAAGLTLRLFRATLADPDGDSNRMAFVS	60
Db	80	FSGHIGVGRADCTGOVADINLMGYSKGNQNAAGLTLRLFRATLADPDGDSNRMAFVS	139
Qy	61	ELCHISQRLRLVLRKLESGYSLYRDNVLSAITHSGPAGFPQYTLXILASEGFSNR	120
Db	140	DIGVNSQRLRLVLRKLESGYSLYRDNVLSAITHSGPAGFPQYTLXILASEGFSNR	199
Qy	121	TFQVYVSGIMKSIDIAHTNKLPGKIFLNKGNVAVQINRSPYSTLQNPQSERARPSMTD	180
Db	200	TFQVYVSGIMKSIDIAHTNKLPGKIFLNKGNVAVQINRSPYSTLQNPQSERARPSMTD	259
Qy	181	KEMVLKLVLDNGDLGLISWFAIHPVSMNNSNHFVSDNMGVAAVLEDEKKNKGYLFGQ	240
Db	260	KEMVLKLVLDNGDLGLISWFAIHPVSMNNSNHFVSDNMGVAAVLEDEKKNKGYLFGQ	319
Qy	241	GPFVAGRASSNLGVSPNLLGPHCVNTGESCDNKSCTCPNGSPCMASGPGQDMFESTH	300
Db	320	GPFVAGRASSNLGVSPNLLGPHCVNTGESCDNKSCTCPNGSPCMASGPGQDMFESTH	379
Qy	301	IIGHIYOKAKELTASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAA	360
Db	380	IIGHIYOKAKELTASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAA	439
Qy	361	GTIDVSGNLITGCTTGGDPFMDLRLQDLGKPESEIEVCOKPKPILHSGELTIPHPWQ	420
Db	440	GTIDVSGNLITGCTTGGDPFMDLRLQDLGKPESEIEVCOKPKPILHSGELTIPHPWQ	499
Qy	421	PDIVDQIVYVGSIAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNYTH	480
Db	500	PDIVDQIVYVGSIAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNYTH	559
Qy	481	YITTEEXQAOQRYEASTIYGPHTLSAYIOLFRDLAKAIAIDTVIANNSSGPEPFKLI	540
Db	560	YITTEEXQAOQRYEASTIYGPHTLSAYIOLFRDLAKAIAIDTVIANNSSGPEPFKLI	619
Qy	541	ASLIPNADRAPICKHGDVLOPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEK	600
Db	620	VPLIPSIYDRAPKGRTEGDVLOPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEK	679
Qy	601	YEDSVADQIMYNDASMETRYWIKGLISNATIIYHIPTAIPGIYIRIFGHNRKOE	660
Db	680	YEDSVADQIMYNDASMETRYWIKGLISNATIIYHIPTAIPGIYIRIFGHNRKOE	739
Qy	661	LLKPAVILAFEGISSPEFVVT 681	
Db	740	LLKPAVILAFEGISSPEFVVT 760	

RESULT 4  
08R236 PRELIMINARY: PRT: 496 AA.

Qy 01-JUN-2002 (Tremblrel. 21, Created)  
Dy 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
Dt 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
De Similar to N-acylsphingosine amidohydrolase 2 (Fragment).  
Os Mus musculus (Mouse).  
Oc Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Ox Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
Nc NCB1\_TaxID=10090;  
Rp SEQUENCE FROM N.A.  
Rc TISSUE=LIVER;  
Ra Strausberg R.;  
Rl Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
Dr EMBL; BC022604; AAH22604.1;  
Kw Hydrolase.  
Ft NON TER  
Sq SEQUENCE 496 AA; 54554 MW; 0206597694D128C6 CRC64;

Query Match 73.5%; Score 2634; DB 11; Length 496;  
Best local similarity 100.0%; Pred. No. 1,le-200;

Qy	187	KLVDNGEDDLGLISWFAIHPVSMNNSNHFVSDNMGVAAVLEDEKKNKGYLFGQPFAG	246
Db	1	KLVDNGEDDLGLISWFAIHPVSMNNSNHFVSDNMGVAAVLEDEKKNKGYLFGQPFAG	60
Qy	247	FASNLGVSPNLLGPHCVNTGESCDNKSCTCPNGSPCMASGPGQDMFESTHIGRI	306
Db	61	FASNLGVSPNLLGPHCVNTGESCDNKSCTCPNGSPCMASGPGQDMFESTHIGRI	120
Qy	307	YOKAKELTASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAAGTIDGV	366
Db	121	YOKAKELTASASQEVTPVLAHQVNMNTDVSVOLNATHVTCTCPALGYSFAAGTIDGV	180
Qy	367	SGNLITGCTTGGDPFMDLRLQDLGKPESEIEVCOKPKPILHSGELTIPHPWQDIDV	426
Db	181	SGNLITGCTTGGDPFMDLRLQDLGKPESEIEVCOKPKPILHSGELTIPHPWQDIDV	240
Qy	427	QIVTVGSIAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNYTHIITYE	486
Db	241	QIVTVGSIAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNYTHIITYE	300
Qy	487	EYQAOQRYEASTIYGPHTLSAYIOLFRDLAKAIAIDTVIANNSSGPEPFKLI	546
Db	301	EYQAOQRYEASTIYGPHTLSAYIOLFRDLAKAIAIDTVIANNSSGPEPFKLI	360
Qy	547	IADRAPICKHGDVLOPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEKEDSA	606
Db	361	IADRAPICKHGDVLOPAKPEYRGEVEVIFVGNAPKNSAENQTHQFLVEKEDSA	420
Qy	607	DMQIMYNDASMETRYWIKGLISNATIIYHIPTAIPGIYIRIFGHNRKOE	666
Db	421	DMQIMYNDASMETRYWIKGLISNATIIYHIPTAIPGIYIRIFGHNRKOE	480
Qy	667	LLAFEGISSPEFVVT 682	
Db	481	LLAFEGISSPEFVVT 496	

Matches 496; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5  
Q9VA70 PRELIMINARY: PRT: 704 AA.

Qy 01-MAY-2000 (Tremblrel. 13, Created)  
Dy 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
Dt 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
De CG1471 protein.  
Os Drosophila melanogaster (Fruit fly).  
Oc Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Ox Ephydroidea; Drosophilidae; Drosophila.  
Nc NCB1\_TaxID=7227;  
Rp SEQUENCE FROM N.A.  
Rc STRAIN=BERKELEY;  
Ra MEDLINE=20196006; PubMed=10731132;  
Rl Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Rl Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
Rl George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Rl Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Rl Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chumpe M., Pfeiffer B.D.,  
Rl Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Rl Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Rl Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Rl Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
Rl Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
Rl Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Rl Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
Rl de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Rl Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Rl Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Rl Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,





Db 323 STDIIIGNOFNKLELEFNNAITOVSGKIQRHTWKPFNTNSVEAPNVSGEGATTCCGAM 382  
 QY 355 GYSFAAGTTIDVSGSLNTTQC-TTEGDFPMDTLRDOLGKPSSEIEVCOKRPILLHSGEL 413  
 Db 383 GYSFAGGTGDPGAFNFIQGDNSTTNPFNFI-GGIATKPTPOATCAQAPKPIIDVG-M 440  
 QY 414 TIRPMOPDIVDQIVTYSGLAIAIPGELTMSGRFRFAIKKEPALYMKDKMTVIAG 473  
 Db 441 VEPIDPMVDPMLQIITLQIIVLAVPGEFTMSGRRLNTV-REIIGSIEINPIVLIAG 499  
 QY 474 LSNVHTYITTEYEOAROEASTYIGPHTLSAYIOLFRLAKAITADTVANSSSP-- 531  
 Db 500 LAMTYSGLTTEFEFOVQREAGSTVFGPHTLGAQOEFPANLAOSTYDGSQADPGTPRN 559  
 QY 532 ---EPPEFKNLIASLIPNIADRAPIGKHFQDVLQPAK--PEYRGEVEYEVFGANPKNS 586  
 Db 560 MSCHTPEF-----LPPVIVDVARPKPDDFDITDVTSTTPSYINQVTVYIFGANNRN 613  
 QY 587 AENQHTQFTLVKRYEDSVADQIMNDASMETRFVHKIGLISNATYIWHIPDTAYPG 646  
 Db 614 F--MTSSSFLTVDOLO-SNGQWTTIINDGDMDTKLWKMHDLGFSLITVDMTISPTPOG 670  
 QY 647 IYIRFYGNRKQELLKPAVILAFEGISSPEFV 679  
 Db 671 TYRITHSGVAKKNPFSDN--LTRYQGISNFMV 701

## RESULT 9

Q9SH86 PRELIMINARY; PRT; 715 AA.  
 AC Q9SH86;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE A2938010 protein.  
 GN A2938010.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_Taxid=3702;  
 RX MEDLINE=20083487; Pubmed=10617197;  
 RC STRAIN-CV. COLUMBIA;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umeyan L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Lin X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC007661; AAD32770.1; -  
 DR InterPro: IPR000418; Ets.  
 DR PROSITE: PS00345; ETS\_DOMAIN\_1; 1.  
 SQ SEQUENCE 715 AA: 78505 MW: 07584099F9E2AEEB CRC64;

Query Match 34.9% Score 1249.5 DB 10; Length 715;

Best Local Similarity 37.4%; Pident. NO. 2.3e-90; Indels 129; Gaps 18;

Matches 280; Conservative 113; Mismatches 226; Indels 129; Gaps 18;

QY 4 YTVGRADCTGVSDINLNGYKNGCONARGLTRLFSAFLADPDGSRMAFVSVELC 63  
 Db 26 YLIGVSYDITGPAADVMMGMVANSQDIASGIFRLARAFIYAE----- 71

QY 64 MISORLEVLKRLKESYSGLYRRDNVILSIHTHSGPAGFYTYLIIASEGFSNTEQ 123  
 Db 72 -----OENVAIGSIHTHAGPGGLQVYTVYISLGFVRSQSD 109  
 QY 124 YIVSGIKSDIAHTNLPKPKIFLNKGNVANNVQINRSPSYLNPSEARARYSSNDKEM 163  
 Db 110 VVANGIEOSTVOAHESLRPSAFVKNKGLDLADAGVNRSPSSYLNNPAERSKYVDKEM 169  
 QY 184 LVKLVDLNEEDLGLISMFAIHPVSNMNSNHPVNSDMGYAAVLFEO-----EKNK---- 234  
 Db 170 TLVKFVDSQGLPGPSFWMFTHTGSMRTSLISGDKGAARFMEDWFGOKNSVSSR 229  
 QY 235 -----GYLPGOG-----PRVAGFAS 250  
 Db 230 NIPRSTVSDSRNOSRLDIAATYKSSRGSVDKSLDVKTRVRNKRKKEVSAFCQS 289  
 QY 251 NLGDVSNILGPHCVNNGESCDDKSPRCNGSGMCAASPG-QDMESHIIIGRIITYOK 309  
 Db 290 NCGDVSPNTLGTFCIDTGLPCDFNHSTC-NGONELCYGRGPGYPDEFESTRIIGEKOFM 348  
 QY 310 AKELYSASQEVTPVLAHQWVNMEDVSVOL-----NATHVTCKPALGYSPAAGTIDG 365  
 Db 349 AVELFNKATEKLGKIGYQHAYIDFSMLDVTYKAGGSGSTVTCRAANGFGAAGTIDG 408  
 QY 366 VSGLNTQGTTEGDPFMDTLRDOLGKPSSEIEVCOKRPILLHSGELTIPHPMPDIVD 425  
 Db 409 PGAFDFKQGDGDNVFWRLVRN-VLRTPGPEQVQCKPKPILLDTGEMKEPYDM----- 461  
 QY 426 VOIVTVGSLAIAIPGELTMSGRFRFAIK-----KEPALYMKDKMTVIAGLSNV 477  
 Db 462 -ALIRIGQVILSPGFEFTMGARLRLDAIKSELISSDPKES-----NNHNVYIAGLINT 516  
 QY 478 YTHYITTEYEQAROEASTYIGPHTLSAYIOLFRLAKAITADTVANSSSGPEPP-FF 536  
 Db 517 YSQYIAFEEYEVQRYGASTLVGRHTLVAYIQEKKALATALVNG--LTPRPGPPDLD 574  
 QY 537 KNLIASLIPNIADRAPIGKHFQVLO--PAKPEYRGEVEYEVFGANPKNSLENQHTQ 594  
 Db 575 DKISILSPVAVVSTPLGVGEKVDKADVPKSTFRROQVNAFTWSCGPRN--DLMEGS 632  
 QY 595 FLTYEKEDSVADQIMNDASMETRFVHK--GIIGLSNATYIWHIPDTAYPGIYIRY 652  
 Db 633 FAYETIREG-GKAPRYDDDDSLKFKMSRPKLSSEQATTEMRPESAAGVYRIRH 691  
 QY 653 FGNHRKQELLKPAVILAFEGISSPEFV 680  
 Db 692 YGASKSLF-----GSISFSGSSSAFYV 715

## RESULT 10

Q9FILA PRELIMINARY; PRT; 705 AA.  
 AC Q9FILA;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE Neutral ceramidase.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NCBI\_Taxid=3702;  
 RX MEDLINE=99156233; Pubmed=10048488;  
 RC STRAIN-COLUMBIA;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned pl and TAC clones.";  
 RL DNA Res. 5:379-391(1998).  
 DR EMBL: AB018885; BAB09641.1; -

SQ SEQUENCE 705 AA; 78426 MW; 52C77CD9F823B680 CRC64;

Query Match	34.1%;	Score 1223;	DB 10;	Length 705;
Best Local Similarity	38.1%;	Pred. NO. 2.9e-88;		
Matches 287;	Conservative 105;	Mismatches 211;	Indels 150;	Gaps 22

QY	2	SGYIIGRADCTQVSDINLMGCGKNQONRGLTLTFEASFIADDDGSNRBAFVSE	61
Db	27	SDYLMGIGSYDITGPADVNMGMAGNNEQVASGVHFLRAAFIAVEP-----	74
QY	62	LCMIQSRLLEVLKRLKESKIGSLYRDNVTLSAIHTHSGPAGFOYTLTLASEGSNRT	121
Db	75	-----YKENVAISGTHHTHAGPGCYLOYILYLTSLSEFVHQS	110
QY	122	FOYIVSGIMKSIDIAHTNLKPKGIPIKKNVAVANOINKSPSSYLINPOSEARVSSNDK	181
Db	111	FNALVDSIGESIIIOAHENLRPGSLINKGELLDGAVNRSPSAYLINPNHESKYEYDVK	170
QY	182	EMLVKLTVDLNGEDGLISWFAHPVSNMNSNHFPVSDNMGYAAYL---FEQEKKNKGYL	237
Db	171	EMTLVYKFEVD-----DQW---GPVGSFNC-----GDNKGTAARIMELMEWERENGCRSV	214
QY	238	PGCGP-----FV	244
Db	215	DVEBPFRVYSIISDPYGEHODLMEMASSLSTGCKTYVTMRSSVARVRK:RRRHADKPRV	274
QY	245	AGEASSNLGADVSPNIIILGPHCVNTGESCNDKSTCGPNCGPSCMASGPG-ODMFESTHTIG	303
Db	275	SACQJTCGCGSPVPLVAFICIDITGPCEFNOST-GGKNQCYGCGPGYPRDEFESTRIIG	333
QY	304	RIIYORKAKELVYASQSEVTGPVYLAHQVMNNTDVSQVLN---ATHVYKTKCPALGYSA	359
Db	334	ERQFKRADLFTKASELQGVDRYRHAIVDSQLVITINQONGSEVYKTCPAAGFGPA	393
QY	360	AGTIDGVSGNTIGTTEGDEPFMDTLRDQLKGKSEELVEECOKRPKILHSGELTIPHPW	419
Db	394	AGTTDGCAGDFPKGDOGNPFMYLVNR-LTKNPTBEQVRCORPRLDITGEMQPYDW	452
QY	420	QPOLVDQVQYTWGSLATAIPGELTMSGRFRRAIK---KEPALYGGKMDVTYIAGLSN	476
Db	453	-----ALLRIGQVLICVPGEEFTYMGRLRAVAKVLKEGS--NQRERSVVIAGLTN	503
QY	477	VYTHYITVEEYOQRYEAASTIGPPTLSAYIOLFR---DLAKAITDVVANSNGSPE	532
Db	504	SYSGYIATFEYQVQRYEGASTIGCPHTLSGYIDQFKLANLDLSAQPTD-----PGPQ	557
QY	533	PP-PEFKULIASLIPNADRAPIGHNEDVLO-PAKPEYRG-EYEVETVPGANKNSAE	588
Db	558	PPOLLKHOISLTPPVADMTPICGAFDVSIDVRLSKFRKGADIVRQVQFSANDRN--D	615
QY	589	NOTHQETLVEKYEDSVADMQIMNDASWETRFYMHK--GILGSLNATVYHIHPTAVPG	646
Db	616	LMTEGFPALVERMTEGRETWPVYVDDDFCLRFKMSRFRKLSSTQSTAIEMRIPEPTASPG	675
QY	647	IYRIRYRGHNRKQDELKPAVILAEFGISSPREV	679
Db	676	VYRITHFG-SAKTPI---SSIHHSFGSSSAFVV	704
RESULT 11			
08X0X9			
ID	08X0X9	PRELIMINARY:	PRT: 841 AA.
AC	01-MAR-2002 (Tremblrel. 20, Created)		
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)		
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)		
DE	Conserved hypothetical protein.		
GN	12344.100.		
OS	Neurospora crassa.		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
OC	Sordariales; Sordariaceae; Neurospora.		
OX	NCBI_Taxid=5141.		
NN	[1]		

RP SEQUENCE FROM N.A.  
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
RA Wakatsuma G., Mewes H.W., Mannhaupt G.,  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA German Neurospora genome project;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL670009; CAD21363.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 841 AA; 92901 MW; 0E3EB570314B3A0B CAC64;

Query Match	31.38;	Score 1120.5;	DB 3;	Length 841;
Best Local Similarity	-36.58;	Pred. No. 5.4e-80;		
Matches 269;	Conservative 119;	Mismatches 279;	Indels 69;	Gaps 23;

ID	Q91596	PRELIMINARY:	PRT:	670 AA.
QY	4	YIIGVGADGCGVSDINIMCYGKNGONAROLLRLTSRAITLADPD--GSRMAFVSVEL	62	
Db	110	YLLCGVGADITGPVVEINLMGYADPKQGLTGRLRLVSRATVSGLEPRDFRYLLVDT	169	
QY	63	CMISQRLR---LEVLAKRLRESKYGSLYLRDNVILSAITHSGSPAGFFOYTLILASEGFSN	119	
Db	170	QSGDTAVRFGIIRAKLKELGREY-AFYGHNNIALTLGTHSHAGPGMLNLLPQITSGGDR	228	
QY	120	RTFOYIYSGIMKSIDIAHTNLKPEKIFLNKGVANVQINSPSYLLNPOSEARY----	175	
Db	229	QGYAIVDGVAVLSIRKAHESLQPGYLSAGTTKVGANINRSLFSLANPEMERKKYFKSE	288	
QY	176	--SSNTQKEMVLKLYDL--NGEDGLISWFAIHHPVSMNNSHHFNSDMDMGAAALFDEOK	232	
Db	289	EDDSVERKDLMLKFORASDCKNIGVLTFFVHGHTSHLGNITVYTGDKKGYAANLFEKSV	348	
QY	233	NGYLPQGPVPGVAFSSNLGDVSPNIGPHCV--NTGESCNDKSTCPNGSPSCMACSP	291	
Db	349	-RGDSSAAEDFVAFGSOANGMDVSPNVLGAACEINSGSOCFSKKNSTGCSQACHGRGP	407	
QY	292	G---QDMFSTHIGRIIYOKARELY--ASASQETGP-VLAHQGVNNTDVSQOLNA	343	
Db	408	AFKVHDGASGCEFTIGRQEPAPQOLYDHLINPSPVRGPPVKAFHTFHDSNFTFLRPD	467	
QY	344	THVTKCKPALGYSEFAGTTIDGVSGNLITQ---GTTGDPFMDTLRLDQLGKPSSEIV	398	
Db	468	GKPARTCPALGYSEFAGTTIDGPGAFDPTQINGENMTM--PIWKVVR-YFIKAPGDEQK	524	
QY	339	ECQPKRPIILHSGSLTTPHQPOPIVQIVYVSSLSIAIAPGELTMSGRRFEALIKKE	458	
Db	525	ACQHPKPIILLDVGELIRPYDMSPIIVVQAFRVQOFALIVSPGATTTAGGRMEKAEVQS	584	
QY	459	FALYGMKDM-----TVVLAGLSNVYTHYITTYEEOAQRYEAASTLYGPT	504	
Db	585	KRILLDDMEQTYGYTGGPSSNPVYVLGGRANITYTHITTEEBEQIQRYEASLTIXGPT	644	
QY	505	LSAVIQLFRLAKAIATDVANMSSGPE--PPEFKNLASLIPNIA--DRADIGHFGDVL	561	
Db	645	LMNVIANTLSTFLPLSEESTSPRAHDEKSPDPDNTNSLSRIFPAVVBDAPLFLFGCGDVL	704	
QY	562	QPAKPE--YRNGEVEVLIPIGANKSNSENOTHTQFLTVEK-YEDSV---ADMQIWNND	614	
Db	705	VVDVDSANGYRGGDIAIRAVFGANRRND--RLESTIAVAERKLFVDQVNPQKSEMKRYASD	762	
QY	615	ASWETRFYMHK--GILGLSNATYIWHIIPDTA-----YPIYIRIRFYGNRRQELLKP	664	
Db	763	EDMSLIYSWMKRKNSVMGTSEVEIEMWETGEETDEWMDKELGCVGRYLRKYYGDSKS--LFG	819	
QY	665	AVILIAFEGISPEEYV 680		
Db	820	GKVOEFGVSGAFTLV 835		



DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Hypothetical protein PA0845 (Ceramide-hydrolysing enzyme precursor).  
 GN PA0845.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NC NCB1\_Taxid=287;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15592 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith J.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Nieuwenhuizen W.F.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RA Okino N.;  
 RT "Molecular cloning, sequencing, and expression of the gene encoding  
 RT alkaline ceramidase from Pseudomonas aeruginosa. Cloning of a  
 RT ceramidase homologue from Mycobacterium tuberculosis.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE004519; AAC04234.1;  
 DR EMBL: AJ315932; CAC67511.1;  
 KM Hypothetical protein: Signal: Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 670 CERAMIDASE.  
 SQ SEQUENCE 670 AA: 73372 MW: 712073EAC9CED287 CRC64;  
 Query Match 30.4%; Score 1090.5; DB 16; Length 670;  
 Best local Similarity 35.9%; Pred. No. 3.6e-78;  
 Matches 250; Conservative 119; Mismatches 253; Indels 75; Gaps 14;  
 QY 4 YIIGVGRADCTGQVSDINLMGYGKNGQNGARGLTLRFSRAFIADPDGSRNRAFYVELC 63  
 DB 29 YRGLGKADITGGAEEVGMGYSSLEOKTAGIHRQMARAFVIEEASGRRLVYNTDLG 88  
 QY 64 MISQRLRELYKRLSEKSGSLYRRDNVILSAITHSGPAGFQYTYLLASEGFSNRTQ 123  
 DB 89 MFOAVHLKYLAKKYPGVYDENNVMLAATHHSGPGGFSHYAMYNLSVLGFOKRTFN 148  
 QY 124 YIISGIMKSIDAHTNLKPGKIFINKGNVANOINSPSSYLINPOSERARVSSNTDKM 183  
 DB 149 AIVDGIYRSIERQAALQGRFLFYGGSELNARNRSLSLHLKNP--DIYGEDGIDPOM 206  
 QY 184 LVKLVDLNGEDGLISWFAIHPVSNMNSHFNVDNMGYAAYLFQDEKNKGYLPGQGP 243  
 DB 207 SVLSFVDANGELGALISWFPVHSTMTNANHLISPNKGYASYHWEHDVSR---KSGF 261  
 QY 244 VAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCENGSPSKMAGSPQDMFESHTIIG 303  
 DB 262 VAAFAQTNGNLSPLN-----NLKPGSGPDNEPDNREIG 297  
 QY 304 RIITYQAKELIYASASQEVTPVLAHQWVMTDVSQVLANAT--HTVKTCKPALGYSFAAG 361  
 DB 298 LRQFAKAYEIAQAOEVEIGELDSRRFVDFTRLPTRPEFTDQPOLCTAAIGTSLAAG 357  
 QY 362 -TIDGVSGLINTGCTEGDPFMDTLRDQLGKRSSEIIVECQKPKILLHSGELTIRHPWQ 420  
 DB 358 SFEDGGPIGLEEG---NNPFLSALGGLTGVPPEQLVCOQAEKTLTADTGNKK--PYPM 413

QY 421 PDIVDQIYTVGSLAIALIPGELTMTSGRRFRPAIKKEPALGKMDTVYIAGLSNVYH 480  
 DB 414 PYVLPIDMEFICGELTGLGAPAEFTVAGVIRRAQASPAEIRH--VVENGYANAYAS 471  
 QY 481 YITTEEYOQAREASTIYGPHLSAYIOLPFDLAKA-----INTDVA-----N 526  
 DB 472 YITTEEYAOELEGSTLGPWTQANVQQLPDMVALERLPLETSALAPDLSCQKN 531  
 QY 527 MSSGPPPEPFENKLIASIPINADRAPIGKFGDVLQPAKPEYRGEVEYIEVGANKPS 586  
 DB 532 FQTVG-----VADDPYIGKSGFDVLQDPRESYIGKTVYAFYTGHPKN- 575  
 QY 587 AENQOTQFLTVKTY-EDSYADQWIMYNDASWETPRYWHKGLIGLSNATIIHIDPTANP 645  
 DB 576 -DURTEKTFLEVYVNIKDKGQPEYATNDMDQYRWMEVGVISASKATISIPGTEP 634  
 QY 646 GYIRYFEGHNRKQELLKPAVILAFGSISSPEFVYT 682  
 DB 635 GHYIRHYGNANKFWTK---ISEIGSTRSEFVLOT 668  
 RESULT 13  
 Q9RHO0 PRELIMINARY; PRT; 670 AA.  
 AC Q9RHO0;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE Alkaline ceramidase.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NC NCB1\_Taxid=287;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20062886; PubMed=10593963;  
 RA Okino N., Ichinose S., Omori A., Imaiya S., Nakamura T., Ito M.;  
 RT "Molecular cloning, sequencing, and expression of the gene encoding  
 RT alkaline ceramidase from Pseudomonas aeruginosa: Cloning of a  
 RT ceramidase homologue from Mycobacterium tuberculosis.";  
 RL J. Biol. Chem. 274:36616-36622(1999).  
 DR EMBL: AB028646; BAA8409.1;  
 SQ SEQUENCE 670 AA: 73275 MW: 3E9FCEB3521AC0 CRC64;  
 Query Match 30.4%; Score 1090.5; DB 2; Length 670;  
 Best local Similarity 35.9%; Pred. No. 8.9e-78;  
 Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;  
 QY 4 YIIGVGRADCTGQVSDINLMGYGKNGQNGARGLTLRFSRAFIADPDGSRNRAFYVELC 63  
 DB 29 YRGLGKADITGGAEEVGMGYSSLEOKTAGIHRQMARAFVIEEASGRRLVYNTDLG 88  
 QY 64 MISQRLRELYKRLSEKSGSLYRRDNVILSAITHSGPAGFQYTYLLASEGFSNRTQ 123  
 DB 89 MFOAVHLKYLAKKYPGVYDENNVMLAATHHSGPGGFSHYAMYNLSVLGFOKRTFN 148  
 QY 124 YIISGIMKSIDAHTNLKPGKIFINKGNVANOINSPSSYLINPOSERARVSSNTDKM 183  
 DB 149 AIVDGIYRSIERQAALQGRFLFYGGSELNARNRSLSLHLKNP--DIYGEDGIDPOM 206  
 QY 184 LVKLVDLNGEDGLISWFAIHPVSNMNSHFNVDNMGYAAYLFQDEKNKGYLPGQGP 243  
 DB 207 SVLSFVDANGELGALISWFPVHSTMTNANHLISPNKGYASYHWEHDVSR---KSGF 261  
 QY 244 VAGFASNLGDVSPNLTGPHCVNTGSCDNKSTCENGSPSKMAGSPQDMFESHTIIG 303  
 DB 262 VAAFAQTNGNLSPLN-----NLKPGSGPDNEPDNREIG 297  
 QY 304 RIITYQAKELIYASASQEVTPVLAHQWVMTDVSQVLANAT--HTVKTCKPALGYSFAAG 361  
 DB 298 LRQFAKAYEIAQAOEVEIGELDSRRFVDFTRLPTRPEFTDQPOLCTAAIGTSLAAG 357  
 QY 362 -TIDGVSGLINTGCTEGDPFMDTLRDQLGKRSSEIIVECQKPKILLHSGELTIRHPWQ 420



```

Db 358 STEGPGPLGLEEG---NNPFLSALGILTCVPPPELVOCOAEKILIAMTKNKK-PYPWT 413
OY 421 PDIVDOVQIVYVGSALIAIPGELITMSGRREPRRAIKKFEALYGMKDMYVLAGLSNYTH 480
Db 414 PTVLPIQWFRIGQLLELGAPEFTVMACVRIIRAAVQAASEAAGLRH--VVENGAMAYAS 471
OY 481 YITTYEEOAOYEASTIYGPHTLSAYIQLFRLAKA-----IATDYVA-----N 526
Db 472 YVTRREFAAOEYEGSSTLYCPWTOAAQOLFVDMANALRERLPVETISAIAPDLSCQMN 531
OY 527 MSSGPEPPEFKNLILASLIPNADRAPIGKHGFDVLOQAPKPEYRGEVVEYTFVGANPKNS 586
Db 532 FQTVGV-----VADDPYIGKSFQDVLOQPRESEYRIDKVTVAFTGHPKN- 575
OY 587 AENOTHOTELVEXY-EDSVADMOQIMYNDASWETRFYHKKGLIGSNATYIWHIPDTXYP 645
Db 576 -DLTKETFELEVNIAGKQKQTPVTATNDMDQYRNERVGISASKATPISMSIPGTEP 634
OY 646 GIVRIYFGHNRKQELLKPAVILAFEGISSPEFVTT 682
Db 635 GHYIRHGNNAKNFWTK---ISELGGSTRSEFVLGT 668

```

## RESULT 14

006769

ID 006769 PRELIMINARY: PRT: 637 AA.

```

AC 006769:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein RV0669C.
GN RV0669C OR MTC1376.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gents J., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998);
DR EMBL: Z95972; CAB09388.1;
DR TubercuList; RV0669C;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 637 AA; 69489 MW; 5E2D915AB1E4FDB7 CRC64;

```

Query Match 24.7%; Score 884.5; DB 16; Length 637;  
 Best Local Similarity 32.3%; Pred. No. 1.9e-61;  
 Matches 221; Conservative 113; Mismatches 291; Indels 59; Gaps 15;

```

OY 6 IGVRADCTGVSPDINLMGCKNGCNAAGLTLRFLRSRAFIAD--PDGSNNMAEFVSLEIC 63
Db 4 VGRGIADITGEAADCGMGYCKSPQRTAGIHQRLRSRAFRVDRDSQDDADALLILVALP 63
OY 64 MISORLRLVLRKLEKYSGLYRDNDVLSAIFHSGPAGFPQYTYLLASGEFSNRTFQ 123
Db 64 LPMQNVNEEVLRRLADLYGDTYSEQNTLITATHTHAGGCGYLYNLTJSGFRPATFA 123
OY 124 YIVGSKMSIDIAHTNLKPGKIFINKGNVAVNOINRSPSSVLLNQSRARVSSYTDKEM 183
Db 124 AIVDGVESVHAHADVAPEVSLSHGELYGASINRSPSAPDRNPADKAPFPKRVDPHT 183

```

```

OY 184 LVKLVDLNGEDGLISWFAIHPYMSNNSNHFVNSDNNGAYAYLEGE-KNKGYLPGGP 242
Db 184 TLVR-IDRGEATGVCIHFHFAHGTSMTRNNHLSIDNKGPAAYHMERIVGADMGAPD 242
OY 243 FVAGASSNLGDSVNNILGPHCVNTGESCDDNDKSTCPNGGFSMCASAPGDMESTHIT 302
Db 243 FIAAFQNTNPGDMSNVNDGP-----LSPEAPPPREFDNTBRT 279
OY 303 GRIYOKAKELAYASQEVTPGLAAHQVMNMTDVSVOLNAT---HTVKTCKPALGYSEA 359
Db 280 GLCOPEDAFPQLSGATPPIGAG-IDARFTYVDLGSVLYGERTPDGEEERTGCRPMFGAGAM 338
OY 360 AGTIDVAGSLNTQGTGDPFMDPLRDOL--LGKPSSEIYECOKPKPILHSGELTIPH 417
Db 339 AGTDEGPGFCHGRQR---NFPWDLSSAMRIANFT---AAQAPKGIWPARLPNNIH 392
OY 418 FMQPDIVQIVYVGSALIAIPGELITMSGRREPRRAIKKFEALYGMKDMYVLAGLSNV 477
Db 393 PFVOEIVPVQVLRIGRLYLIGIPGPTIVAGIRLRMYA---SIVGADLADVLCGYTNA 449
OY 478 YTHYITTYEEOAOYEASTIYGPHTLSAYIQLFRLAKAIAITDTVANMSSGPEPPEFK 537
Db 450 YIHVYTPPEETLEQRYEGSSTLYCPWTOAAQOLFVDMANALRERLPVETISAIAPDLSCQMN 531
OY 538 NLIASLIPNADRAPI-GKHGFDVLOQAPKPEYRGEVVEYTFVGANPKNSAENOTHOTEL 596
Db 510 SWVRG-----APADAGSFGAVIAEBSATYRPGQAAEAVFVSLPNN--DLRRGTYL 559
OY 597 TVEKYEDSVADMOQIMYNDASWETRFYHKKGLIGSNATYIWHIPDTAVPGIYRIYFGHN 656
Db 560 EVYRREG--ASWVRIADGDWATSFRRMORGRAGSHVISIRMDVPDGTTPGQYRIYHGSTA 617
OY 657 RKQELLKPAVILAFEGISSPEFV 680
Db 618 RDRN---GMLTASATTFEFTV 637

```

## RESULT 15

093216

ID 093216 PRELIMINARY: PRT: 314 AA.

```

AC 093216:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC Chen K., Chen H., Kim C.J., Koeseema E., Meyers M.C., Banh J.,
RC Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamuya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsids cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY057506; AAL09747.1;
SQ SEQUENCE 314 AA; 35401 MW; 6FE09DA6BEFF2434 CRC64;

```

Query Match 17.5%; Score 626; DB 10; Length 314;  
 Best Local Similarity 45.6%; Pred. No. 2.1e-41;  
 Matches 144; Conservative 48; Mismatches 96; Indels 28; Gaps 12;

```

OY 377 EGDPPMDLRLROLLGKPSSEIYECOKPKPILHSGELTIPHPQPDIVQIVYVGSAL 436
Db 13 KNPFWRLVRN-LKNPTEDVRCORPKPILDTGEMKOPYDWAAPSILPVQILRIGQLVI 71

```

```

OY 437 AAIPELTMSGRFRREAIK--KEPALTGMKDMVINGLSNVTHYITTEEYOQR 493
DB 72 LCVPGFTTMAGRRLDAVKYTKES--NGREFSVIAGLTNSOYIATEEYOQRY 129
OY 494 EAASTIYGPHTLSAVYQLFR---DLAKAIATDTVANMSSGPEPP-FFKNLIASLIPNIA 548
DB 130 EGASTIYGPHTLSGYIOEFKKLANDLLSQOTD-----PGPOPDLHKQISLLTPVVA 183
OY 549 DRAPIGKHGDLQ--PAKPEYRVG-EVEVEIVGANPKNSAENQTHQTFITVEKYEDSV 605
DB 184 DMTPIGTAGDYTSVPRLSKFRKGADIVYQFRSANPRN--DLMEGTFFALVERMLEGR 241
OY 606 ADWQIMYNDAWNETRFRYHK--GILGLSNATYIWHIPDTAYPGIYRIRIFGHNRKQELK 663
DB 242 ETWVPVYDDDDCLRFKMSRPFKLSTQSTATIEMRIPETASPGVYRITHFG-SAKTPI-- 298
OY 664 PAVILAFEGISSPPEY 679
DB 299 -SSIHFGSGSSAFV 313

```

Search completed: July 3, 2003, 12:47:33  
 Job time : 95 secs



Db 335 KRC---FVSAFCOTNCGDVSPNVLGAFCLDTGLPCPNHSTC--CGKNEMCYGRGPGPDE 390  
QY 296 FESHTIIIGRIYOKAKELASASOEVTGPVLAHOMVNMIDVSVOL-----NATHVTCK 351  
Db 391 FESTRIIGEHOEFMALELEFNKASEQLOGVKDYRHVYVDSQLVTLPKKCKGSEVYKTCR 450  
QY 352 PALGYSFAGTIDVSGCLNITOGTTGDDPFWTLRDQLGKPSSEIVECOCKPRPILHSG 411  
Db 451 AAMGEFAAGTDDGPAFPTQDDKGNPFWRJVRN-VLKTPOKOIDCHYKPIILDTG 509  
QY 412 ELTIHPMOPDIVDVOIVMGLAIAIGELTMSGRREARAIKKEFALYGMKDMT--- 468  
Db 510 EMKRPIDMAPSILSDVLRIGQLFILSVGEFTTMGRRLRVAKTQOLKNSGKNDJSGEI 569  
QY 469 -VVIAGLSNYTHYTTTVEEYQARYEAATSTYGPHTLSAYIOLFRDLAKAATDVTANM 527  
Db 570 HVIYIAGLANGSOYVTFEEYQVREGASTLYGPHTLSCYIOEFKLSLVLDD--MPV 627  
QY 528 SSGPEPP-FFKNLIASLIRIADRIAGKHFGVLO--PAKPEYRGE-VVEYIEGANG 583  
Db 628 QPQPQPPDLLDKQLSLTPVMDTPSGDSFGDIVSDVPKLSLKGNGQVTVVFRSACP 687  
QY 584 KNSAEQHTQTFLTVEKEEDSVADMOIMYNDASMETREYV--HGILGSLNATIYHHPD 641  
Db 688 RN--DLTGETFLVLERLQKDKTWTPTVDDDLCLRFKSRKHKLSNSQATVENRIPE 745  
QY 642 TAVPGYIRIRYEGHNK-----QELLKPAVILAEFGISSPFEV 679  
Db 746 SASPGYRITHFGCAKKLFQGRVQNGVKELMLRSLMESNIPFHV 789

## RESULT 2

H84799  
hypothetical protein AC2938010 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: H84799  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shua, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
Gues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:1061197  
A:Accession: H84799  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-715 <STO>  
A:Cross-references: GB:AE002093; NID:g4895183; PIDN:AAD32770.1; GSPDB:GN00139  
A:Gene: AC2938010  
A:Map position: 2  
C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV0669c

Query Match 34.9%; Score 1249.5; DB 2: Length 715;  
Best Local Similarity 37.4%; Pred. No. 9.9e-85;  
Matches 280; Conservative 113; Mismatches 226; Indels 129; Gaps 18;  
QY 4 YIYVGRADCTGVSDINIMGYKNGONARGLTLRLFSRAFLIADPDGSRMAFVSELC 63  
Db 26 YIIGVGSYDTGTGAADVNMGYANSQDIASGIFRLRARAFIYAE----- 71  
QY 64 MISQRLREYLKRLSKYSGSLRRDNYILSAITHSGPAGFOYTYLIASGFSNRTQ 123  
Db 72 -----OEKNVAISGITHAGPGGYLYQYTYVYISLGFVRSQSF 109  
QY 124 YIVSGIMKSIDIAHTNLKPKGKIFINKGNVANOINSPSSYLLNPOSERARYSSNDKEM 183  
Db 110 VVYVNGEISIVQAHESLRPGSAFVNGDLDAGVNSPSSYLLNPAERSKYTYVDKEM 169  
QY 184 LVLLKLDLNGEDLGLISWFAIHPVSNMNSNHFNVDNMGYAALFQO-----EKNK----- 234  
Db 170 TLVKKFVDSGLGPTGSEFMWATHTGTSKRSINSLISGKNKAAAFMDFENGNKNSVSR 229

QY 235 -----GYLPGQG-----PVPAGPASS 250  
Db 230 NIPRRVSTIVDSERNQSRLLDIAATYKSSRGHSVDKSLDVKTRVNGSKRKFVSAFCS 289  
QY 251 NIGDVSBNLIGPRCHVMTGESCNDKSTCPNGGSMCMASGPG-QDMFESHIIIGRIYCK 309  
Db 290 NCGDVSPNLTGTICIDTGLPCDPNHSCT--NGONELCYGRPGYPDEFESTRILIGKOFM 348  
QY 310 AKELYSASQEVGPVLAHOMVNMIDVSVOL-----NATHVTCKPALGYSFAAGTIDG 365  
Db 349 AVELFKATPEKLOGKIGYQIAYLDFSNLDVTYKAGGSETVATCAANGFGFAACTTIG 408  
QY 366 VSLNITOGTTGDDPFWTLRDQLGKPSSEIVECOCKPRPILHSELTIHPMOPDIVD 425  
Db 409 PGAFEDKQGGDDGNVFWRLVRN-VLRTPGPEYQOCCPKPPIILDTGEMKEPYDM----- 461  
QY 426 VOIVTIGSLAIAIPBELTMSGRREARAIK-----KEFALYGMKDMTIVYIAGLSNY 477  
Db 462 -ALIRIGOLVILSVPGFTTMAGRRLDLAKLSFLISSDKERS-----NNHVIYAGLITN 516  
QY 478 YTHYITTYEYQARYEAATSTYGPHTLSAYIOLFRDLAKAATDVTANMSSGPEPP-FF 536  
Db 517 YSQYIATPEEYEQVREGASTLYGRHTLPAVIOEFKKTALVNG--LTLPRGQPPDLL 574  
QY 537 KNLIASLIRIADRIAGKHFGVLO--PAKPEYRGEVYEVFVANGPKNSAENDTROT 594  
Db 575 DKQISLSPVYVDSITLGVYFGVADVPKSTFRGQOVNATFWSGCPRN--DLMTES 632  
QY 595 FLTVKEEDSVADMOIMYNDASMETREYVHK--GIILGSLNATIYHHPYAGIYRIR 652  
Db 633 FAVETLRGCG-KGMAVYVDDDFSLFKSKRPKLSSESQATTEMKVPESAAGVIRIRH 691  
QY 653 FGHNRKQELKPAVILAEFGISSPFEV 680  
Db 692 YGASKSLF-----GSISFSSSSSAFYV 715

## RESULT 3

C83540  
conserved hypothetical protein PA0845 [imported] - Pseudomonas aeruginosa (strain PAO  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83540  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82850; MUID:20437337; PMID:10984043  
A:Accession: C83540  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-670 <STO>  
A:Cross-references: GB:AE004519; GB:AE004091; NID:g9946736; PIDN:AAG04234.1; GSPDB:GN  
A:Experimental source: strain PA01  
A:Gene: PA0845  
C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV0669c

Query Match 30.6%; Score 1095.5; DB 2: Length 670;  
Best Local Similarity 35.9%; Pred. No. 2.6e-73;  
Matches 250; Conservative 119; Mismatches 253; Indels 75; Gaps 14;  
QY 4 YIYVGRADCTGVSDINIMGYKNGONARGLTLRLFSRAFLIADPDGSRMAFVSELC 63  
Db 29 YRGLCKADITGTGAABVGMGYSLDQKTAGIMRQMARAFIYEAAASGRLLVYVTDLG 88  
QY 64 MISQRLREYLKRLSKYSGSLRRDNYILSAITHSGPAGFOYTYLIASGFSNRTQ 123  
Db 89 MIFQAVHLKVLARLAKRYGCVYDENNVMLAATHSGPGGFSHYAMYNLSVLFQFKETPN 148  
QY 124 YIYSGIMKSIDIAHTNLKPKGKIFINKGNVANOINSPSSYLLNPOSERARYSSNDKEM 183

Db 149 AIVGVISIRIAQAARLPGRLLFYSGGELRANRRSLLSHLKNP--LIVGDEIDIDPM 206  
184 LVLKLVNDGEGDLGLISWFAIHPIVSMNSNFVSDNNGYAALFEQKKNKGYLPGCGP 243  
Db 207 SVLSFVDANGELAGISWFPVSHSTMTNANHLISDNKGYASYHHEHVSF---KSGF 261  
Qy 244 VAGFASNLGDVSPNLLGPHCVNTGESCDNDKSTCPNGPSCMASGPGQDMFESTHII 303  
Db 262 VAAPAQTNAGLSRNL-----NLKPSGPFDEFDNTRIG 297  
Qy 304 RIIVOKAKELIASASQEVTPVLAHQMVMNTDVSVOLNAT--HTVTKCKPALGYSPAAG 361  
Db 298 LROPKAAVEIAGQAOEVLGELDSRFVDEFTLRPIREFTDGPQRCQTAIGSLAAG 357  
Qy 362 -TIDVSGSLNTGGTEDEDPFMDLRDOLLKPSSEIVECOKPKPILHSGELTIPHQ 420  
Db 358 STEGPGPLGLEEG---NNPFLSALGGLLTGVPPELVQCOAEKTLADTGNKK--PYEWT 413  
Qy 421 PDIVDVOIVTVGSLAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNVYTH 480  
Db 414 PTVLPIDQFRIGQLELLGAPREFYVMAVRIRRAVQASAAAGLIH--VFNGTANAYAS 471  
Qy 481 YITVEEXQAQRYEASTIYSPHTLSAYTQLEFRDLAKA-----IATDTVA-----N 526  
Db 472 YVTREEXAAQYEGSGSTLYGPMWQAAYQQLFVDMAVLRLERLPEVTSALPDLSCCOMN 531  
Qy 527 MSSGPPEPFENLNLASLIPNADRAPKRGKVDVLOPAKPEYRGEVEVYFVGANPENS 586  
Db 532 FQTV-----VADDPYIGKSGVDVLOOPRESYRIGDKTVAFVGHGPKN- 575  
Qy 587 AENOTHOTFLTEVRY-EDSVADMOIMYNDASMETREYVHKGLISNATIYVHPTDAYP 645  
Db 576 -DLRTEKTFLEVVNIGDKQKTPETVATDNDMDYRKREYGISASATISMSLPTGER 634  
Qy 646 GYIRIRYFGNKKQELLKPAVILAFEGISSPEFVTT 682  
Db 635 GHYIRHYGNKAKNFWTK---ISEIGSTRSFEVLGT 668

## RESULT 4

H70535  
hypothetical protein RV0669c - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70535  
R: Cole, S.T.; Brosch, P.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S. Nature 393, 537-544, 1998  
A: Authors: Squares, S.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A: Reference number: A70500; MUID:96295987; PMID:9634230  
A: Accession: H70535  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-637 <COL>  
A: Cross-references: GB:295972; GB:AL123456; NID:93261790; PIDN: CAB09388.1; PID: g2143290  
A: Experimental source: strain H37Rv  
C: Genetics:  
A: Gene: RV0669c  
C: Superfamily: Mycobacterium tuberculosis hypothetical protein RV0669c

Query Match 24.7%; Score 884.5; DB 2; Length 637;  
Best Local Similarity 32.3%; Pred. No. 1.2e-57;  
Matches 221; Conservative 113; Mismatches 291; Indels 59; Gaps 15;

Qy 6 IGVRADCTGVSDINILMGKNGONARGLTFLRSRAFLAD--PDGSNMAFVSELC 63  
Db 4 VGRGIDATGGAADCGMIGYKSDQRTAGIHQRLSRKAFVRDSDADARLLIYALP 63  
Qy 64 MISQRLVLRLESKYGLYRDNVLSAIIHSHGPAFFOYTYLTIASEGFSNRTPQ 123

Db 64 LPMQVNEEVLRLADLYGDIYSQONTLITATHTHAGBGYCYLLNLTSSGRPATFA 123  
Qy 124 YIVSGIMKSIDAHNLNPKGIKFIKKGVANVOIRSSSVLLNQSRFATSSYTKDEM 183  
Db 124 AIVDGVSEVHADVAIPAEVLSHGLGASINRSAPDRNPAPAKAFPPKRVDPHT 183  
Qy 184 LVLKLVNDGEGDLGLISWFAIHPIVSMNSNFVSDNNGYAALFEQKKNKGYLPGCGP 242  
Db 184 TLVR-IDRGEATVGIHFATRHGSMTRNHLISDNKGFAYIHERVGVGADYLAQPD 242  
Qy 243 FVAGFASNLGDVSPNLLGPHCVNTGESCDNDKSTCPNGPSCMASGPGQDMFESTHII 302  
Db 243 FLAFAQYQNPDDMSFNVDP-----LSPEAPPDREEDNTRT 279  
Qy 303 GRIVYOKAKELIASASQEVTPVLAHQMVMNTDVSVOLNAT--HTVTKCKPALGYSPA 359  
Db 280 GLCOFEAFQTLGATPIGAG-IDARFTYVDLSVLYGVEYTPDGEERTSRPMEGAGAM 338  
Qy 360 AGTIDVAGSLNTGGTEDEDPFMDLRDOL--LGRSEIVECOKPKPILHSGELTIPH 417  
Db 339 AGTDGPGPFHGRGCR---NPPMDRLSMTAMRLRPT---AAQAQPKIVPAPRLPNRIH 392  
Qy 418 PMQPDIVDQIVTVGSLAIAIPGELTMSGRPREAIKKEFALYGMKDMTVIAGLSNV 477  
Db 393 PFVQEVIVPQIVRIGRLYLIGIPGPTIVAGRLRBMVA---STVGADLADVLGVYNA 449  
Qy 478 YTHITTYEEYQAQRYEASTIYSPHTLSAYTQLEFRDLAKAIDTVANMSSGPPEPFK 537  
Db 450 YIHVYTPPEEYLEORYEGSSTLYGKMELCALMOTVLAELAMRGPRVTLGRPPREL 509  
Qy 538 NLIASLIPNIDRAP-I-GKHFQDYQAPKPEYRGEVEVYFVGANPENSANOTHOTFL 596  
Db 510 SWVRG-----APADGSGFAGVIAEASATYRQGAIVEAFVSLPN--DLRGGTYL 559  
Qy 597 TVEKEDSVADMOIMYNDASMETREYVHKGLISNATIYVHPTDAYPIYRLRYFGN 656  
Db 560 EVVRREG--ASWVRADDDGDMATSPRMORGRAGSHVIRMDVDPDTPGQYRIVHNGTA 617  
Qy 657 RKQELLKPAVILAFEGISSPEFVY 680  
Db 618 RDRN---GMLTASATREFTVV 637

## RESULT 5

A11489  
Probable peptidoglycan bound protein (LPXMG motif) lin0457 [imported] - Listeria inno  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: A11489  
R: Glasner, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A: Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
Ok, C.; Schuete, T.; Simoes, N.; Tierree, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A: Title: Comparative genomics of Listeria species.  
A: Reference number: AB1077; MUID:21537279; PMID:11679669  
A: Accession: A11489  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-2013 <GLA>  
A: Cross-references: GB:AL592022; PIDN: CAC95689.1; PID: g16412898; GSPDB: GN00178  
A: Experimental source: strain C1p11262  
C: Genetics:  
A: Gene: lin0457

Query Match 3.7%; Score 131.5; DB 2; Length 2013;  
Best Local Similarity 19.8%; Pred. No. 0.92;  
Matches 147; Conservative 95; Mismatches 266; Indels 235; Gaps 37;

Qy 44 FLADPDGSNRMAF-----VSEVLCMISQRLRLEVKLRLESYGLR--RDNVILSAI 95  
Db 102 YLASQTSKYELEKEDGVQAKIDLIVKNDVEVAKTVERSNQLLRSSISDRLFLQA- 160

```

QY 96 HHSGPAGFFOYTLVLLASEGFSNRTFOYIVSGIMKSIDIAHTNLPGKIFINKGVANY 155
D 161 --DSSKATLANYTEQI-----TFNSINFLDSS-----TTLNKGKLVIDFN--SNL 203
QY 156 QINRSSSSYLLNQSERARYSSNTDKEMLVLYVD--LNGE--DLGLI--SMFAIRPVSM 209
D 204 ELVNYPKDYPAANNINISTYSALTGK--LTINLVNIDISSGAPPDIPVAVAGARGVP 261
QY 210 NNSNHFVNSDNMGYAY-----LFEQEKKNKCYLLPGQGPVAG-----246
D 262 MNLKALSGENSGATYTAETKTYVNLNENSSNDY-----SPITGDNNSMARMKELAR 317
QY 247 -----FASSNLGD--VSPNIILOPHCVNTGESCDNDK 275
D 318 SLKPGGYTIOWPEIQKSSBNKFKMLKLEFLKENGDDISVNTADPYVIRFGEPIWSQL 377
QY 276 STCPNGSPSCMASGCGDMFE-----STHIGRIITYOKAKELYASA--SEVYGPVLAH 329
D 378 STV--NGKANVYNDEKQVVEGPINANTYORIQVSMAPKIPADAVKGTETGYVNYD 435
QY 330 QWVNMVDSVOLNATHVTKCKPALGYSPAAGTIDG--VSGLNTTGTTEGDPFMDLRDQ 388
D 436 EDLVNTSIRIKTEVTDSATSI-----AVDSKVSSTISSEGDYLEMGMPRISSA 464
QY 389 LIGKPSSEIVECQKPKPILL-----HSGELTTPHP-----WOPDIY 424
D 485 AFGVNDLEIV-APIPKGIKVLSTYIPNNNSMASMKLEYONGKVASMAPQTSGMPFSKI 543
QY 425 DVOIYVGSIAIAIRIGELT-----TMSGRFRAIKKEKALKGMKD 466
D 544 DQSVNRIEKRLKLSRGIILNDKMDPYTHGTIRMONTGVKAGESF--TLQPESTITDSD 601
QY 467 MFVVIAGLSNVYHYTYTTEEYOQRYEAST-----YGPHTLSAVIOL 511
D 602 KT-----SKAIDTMTASVEK--KVQVVEKSTPRAKINGDVELSTAGIYCKGESEI--I 651
QY 512 FRDLAKAINTDVANNSGDEPPEFKNLASLIPNADRAPIGKHGVDYLQPAKPEYRVG 571
D 652 FFGNDKIAQSVRLSGYSKLENPY-----IFVVVPKCIDVETM--KNP---YR-- 698
QY 572 EYVEYFVGVAN-----PKNSAENQTHQFTLYEKEDSVADQWQIYNDAWMTREYVHK 625
D 699 SIINTYVAPANGNTNLYPKSSAD-----VKKRETLSDGSTLY-----735
QY 626 GILGLSNATYWHIPDTAY-PGI 647
D 736 -----YMEAPDTGLAPGM 748

```

## RESULT 6

```

AF0263
protease IV (EC 3.4.21.-) [Imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF0263
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; M01D:214/0413; PMID:11586360
A:Accession: AF0263
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-616 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90970.1; PID:g15980166; GSPDB:GN00175
C:Genetics:
A:Gene: sppA
C:Superfamily: protease IV
C:Keywords: hydrolase; serine protease

```

Query Match

3.7%; Score 131; DB 2; Length 616;

```

Best Local Similarity 20.7%; Pred. No. 0.15;
Matches 156; Conservative 87; Mismatches 223; Indels 210; Gaps 33;
QY 102 AGFOYTLVLLASEGFSNRTFOYIVSGIMKSIDIAHTNL-----KP-----GKIFINKG 150
D 9 AGFFKMTWRL-----NFTREILNLFALLILIGVYIFQROSPVPEVKALLVNS 62
QY 151 NVANVO--IN--RSPSSYLLNPOSERARYSS-----MTDREMLVLYVDIN 192
D 63 GVIYDPAIANNKLRQNGRELLGASSNRLQNSLFDIYETIRLAKDDDNINGVLISDILT 122
QY 193 GEGLGISWF--AIHFVSNMNSNHFVNSDNMGAAALFEDEKKGKYLPGQGP--VAGFAS 249
D 123 GADQSSLYQYIGKALREFRDTGKRIYVGSYNQTOYLLASFARKIYLSPOGAVDLHGFAS 182
QY 250 SNLDGVSPIILGPHCVNTGESCDNDKSTCPNGSPSCMASGPGDMFESETHIIGRIITYOK 309
D 183 NNL-----YYSLEN-----LKVTTNIFRVCTYS 208
QY 310 AKEL-----YASAOEYTGPIVLAHQVN-----MTDVSVOLNATHVTKCKPALGYSF 358
D 209 AVEPMIRNDKSAARE-----ADSRVGGLMQNYLTIVSANRLT-----PEQLFG 255
QY 359 AAGTIDGSGLNTTGTTEGDPFMDLRDQLCKPSEE--IYVC-----400
D 256 AAGVI--SGLOVAGSGQAKYALDSKLYDOLARPEVASLVAFGMNKKTNDENYISY 312
QY 401 -OKRP-----ILLHSGELTTPHPWOPDIYDVOIYVGSIAIA-----438
D 313 DYOPTAPPOGEOIYVLFANGAL-IDGPPRG-----NVGDDTLAAQIQRALDPKIK 364
QY 439 -----IPGLTMSGRFRRAIKKEKALY--GKMDYVVIAGLSNVYHYTYTTEEYO 489
D 365 AVILRVNSPQGSVAS-----ELIRAEALRAHNPVLSMGMAAGGYWISTPANV- 418
QY 490 AQRYEASTYIGPHTLSAVIOLF-----ROLAK-----AIATPTVNMNS--SGDEPPE 535
D 419 -----IVASPTLTGSGIFGVINTFONSASIGVHTDGVATSPLDVSLTKALPEE 470
QY 536 FKNLIASLIPN-----IADRAPIGKHGVDYLQPAKPEYRGEVE--YIFGANPK--NSAEN 589
D 471 FSGMQMINIENGKTFIDLVATSRH-----KTEQYDQIQAQGHVWIGLAKSGLYD 522
QY 590 QTHQFTLYEKEDSVADQWQIYNDAWMTREYVHKG-----ILGLSNATYWHIP 640
D 523 Q-----LQDFEDAVKKAELAKLTQWLMFVDEPSLDTLIGMSASVHAMLP 571

```

## RESULT 7

```

AD1129
probable peptidoglycan bound protein (LPXTG motif) lmo0435 [Imported] - Listeria mono
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1129
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baguerio, F.; Berche, P.; Bloe
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussuguet, O.; Entian, K.D.; Fsihl,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; M01D:21357279; PMID:11679669
A:Accession: AD1129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2013 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC98514.1; PID:q16409812; GSPDB:GN00177
A:Experimental source: strain Egd-e
C:Genetics:
A:Gene: lmo0435

```

Query Match

3.3%; Score 118.5; DB 2; Length 2013;

Best Local Similarity 19.8%; Pred. No. 8.5;

Matches 148; Conservative 88; Mismatches 266; Indels 247; Gaps 38;

44 FILADPOGSMNAF-----VSELCMIISORLRLKLEKSKYSIVR---RDVNIISAI 95  
102 YLLASQYKYLELEKEDKQAKIDLITKNVDEEAKVERSNROLLSSISDKLFLOA- 160  
96 HTHSGPAGFFQYLYLILASEGFSNRTFOYIVSGIMKSIDIAHTNLKPKKIFINKGNVNV 155  
161 --DSKATLANTYEQI-----TFNYSINFLDGS-----TNLKNKNIWIDERN-SNL 203  
156 QINRSPSSYLLNPOSEARVSSNTDKEMVLKLV-DLGE--DWALHPVSM 209  
204 ELVYPRKDTANRNRIKILYSALTK--LTINLVNISSGAPFDIPYIVLAGYAKPCLP 261  
210 NNSHFNPSDMNGAAY-----LFEQKKKGYLPQGGPFVAG----- 246  
262 MNLKATLSGENSGGTYTPSEKTTTVNLESSSNODY---SPITAGCNNAFNKELSY 317  
247 -----FASNLGDV-SPNIGPHCVNTGE----- 269  
318 SLKPGYTIQWPELOKKSLENKSPKNLKLEVLKNGGVISVNTADPIVIRFGEFWSQL 377  
270 SCQNDK-STCPNGSPSCMASGPGQDMESTHIIIRIYOKAKELVAS-----ASQEV 321  
378 STVGKKNVILNDEKQVVEYGP-----INANIYOKIQVMAKIPADGEGTEY 427  
322 TGPVLAHOWNMIDVSQVLAHTVKTCKPALGYSPAGTIDG-VSGINITOGTGEDP 380  
428 TGIYVYVDEDEYFITSIKI-----KAEVADSATSIADVSKVSTI---SEGDI 472  
381 F-WDLTRDQLLGKPESEIVECOKRP-----ILHSGELTIPHP 418  
473 FEMGFMPRVSSAAGVNDLIVAPIPEGIKALSYIPNNNSMASIKLELYONGKRVSMAP 532  
419 -----WQPDIVDQIVVGSALAIAPG-----ELYTSGRFRRAI 455  
533 QTSQWPFESKIDQSVNRLEKILKTSRQIINDKMPYTHGTIRMQNIQVYKAGESEF-TL 590  
456 KKEPALKGMKDMT-VVLAGLSNVYTHYITTEYEQAOQVEAASITYGHTLSAVY----- 509  
591 RPESTITDPDKTSKIDITTTNSYGNQVVE-----KTSAPAKINGVFLVSTAGISGK 645  
510 -----QLPFDLAKAATDVTANMSSGPEPFKNLIASLIPNADRAPGKHFGVLOPAK 565  
646 GFGSTIFNGDKIKQSVRLSGYSGKLEMPY-----IFVYPRKIDLETP-RNF-IQCP-- 696  
566 PEYVGEVVEYEVGAN-----PKNSAENQHTQFLTEKEYEDSVALLQIMYNDASMET 619  
697 --YR-STLNTYTPAPANGTNTLYPKSSAD-----VKKETLSIGSTLY----- 735  
620 RFVWHKGLIGLSNATITVWHIPDPAV-PCI 647  
736 -----YWEAPDTGLAPGM 748

Db

RESULT 8  
S70843  
hemolytic protein hnda precursor - Haemophilus ducreyi  
N:Alternate names: cytotoxin  
C:Species: Haemophilus ducreyi  
C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 08-Oct-1999  
R:Palmer, K.L.; Munson Jr., R.S.  
Mol. Microbiol. 18, 821-830, 1995  
A:Title: Cloning and characterization of the genes encoding the haemolysin of Haemophilus  
A:Reference number: S70843; MUID:96422469; PMID:8825086  
A:Accession: S70843  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1175 <PAL>  
A:Cross-references: EMBL:U32175; NID:9151070; PIDN:AMC43538.1; PID:9151072  
C:Genetics:  
A:Gene: hnda

F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-1175/Product: hemolytic protein hnda #status predicted <MAT>

Query Match 3.2%; Score 113.5; DB 2; Length 1175;  
Best Local Similarity 15.5%; Pred. No. 8.5;  
Matches 103; Conservative 120; Mismatches 252; Indels 189; Gaps 22;

8 VGRADCTGVSDINLMGCKGNQARGL-LRLFSRAFIADP-----DGSNR 54  
116 LKQELIADYIIV--NPGMSQDGGCFINISASLVGNPITDGLKGRIDGSHN 173  
55 MAFVSELCMIISORLRLKLEKSKYSIVRDNVILSAHTHSGPAGFFQYLYLAS 114  
174 ---ISTTDIAEKTNLILAPVYNIKINIKINGKOVNIIITHNEIN-VNNOQLSISVLPN 229  
115 EG-----FSNRIFQYVSGIMK---SIDIAHTNLKPKKIFINKGN 151  
230 KGKVLGDKIAGSIQANRIRIHTDNRAFLDIETGQKGTNSVYIGAGMLKV-----HGH 283  
152 VANQINRSPSSYLLNPOSEARVSSNTDKEM-----LVKLVLNLDGEGLI----- 199  
284 IDTKNNNNKNNKYYLAKSTQEQASMTGDDINLNTNQLITMAANLTKNNIVLKAASQ 343  
200 -----SWFALHPVSMNNSNHFVNS---DNMGYAAVLFEOEKNKGYL 237  
344 FTEKTLNKHYEKNRRSSGSMYHNNETSKEEINVTYSIKGDNVIVA-----DQAI 396  
238 PGQGFVAGFASNLGDSVSNILGPHCVNTGESCNDKSTCPNGSPSCMASGPGQDMFE 297  
397 DGOSLKITAQNTTTLGKQGVTLRGQINRNREKEKFEFNETSN----- 439  
298 STHIIIRIYOKAKELVASQEVTPVLAHOWNMIDVSQVLAHTVKTCKPALGS 357  
440 --LTGKKFKNQOSQYIASELDIKNLKIGGCIKSGVYRFGDPLV-----N 489  
358 FAAGTIDVSGNLITOGTGEDPFWDLRLDQLGKPESEIVECQ----- 401  
490 TGSTVNAENILN-SNIDNENYWGIGGNSYTKKNSITLQOCADLVINGKYIDSEK 548  
402 -----KPKPILHSGELTIPHPQPDIVDQIVVGSALAIAPGELTMSGR 449  
549 GVKISGRVLSGKEALVRAHKRGLVIDV-KNEITSLVIRKGVY-----DITKASAK 601  
450 RPRE-----AIKKEPFL-----YGMKDMYVYAGLSNYTHYITTE 486  
602 DFKNSVSRGSTVSESMLKLVSDAGDIEVAGSLYKSLGDLVKTDXSNLYVKGQONIM 661  
487 EYQQRVEASTIYGPHTLSAYIQFLRLAKAATDVTANMSSGPEPFKNLIASLIPN 546  
662 ERESEK--SGLTTKG-----NLDPSLIGDITVKNILAVSVTGEVLPDLMQKIDN 707  
547 IADR 550  
708 LQEK 711

Db

RESULT 9  
B81989  
hypothetical protein NMA0688 [imported] - Neisseria meningitidis (strain Z2491 serogr  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holtroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: B81775; MUID:20222556; PMID:10761919  
A:Accession: B81989  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2015 <PAR>  
A:Cross-references: GB:AL162753; GB:AL157959; NID:97379120; PIDN:CAB83974.1; PID:9737  
A:Experimental source: serogroup A, strain Z2491

C:Genetics:  
A:Gene: NMA0688

Query Match  
Best Local Similarity 18.98; Pred. No. 22;  
Matches 111; Conservative 86; Mismatches 204; Indels 272; Gaps 32;

```
3.28; Score 113; DB 2; Length 2015;
109 LYLASEGFSNRTFOYIVSGIMKSIDIAHTNLR-PGKIFLNKGNVANVOINRSPSSYLLN 167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
111 LALLADNNTAKT-----TLLNTPGNLVYHTGKDLNLVADKDLASAIIH 554
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
168 POSERARYSNTDKEMLVKLVNDGEDGLISWFAIHVPVSMNSNHFVSDMGR----- 223
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
555 LKSDNAHITGTSTLTASK--DM-GEVAGLL-----NVTNLTNRNSGLHTQAAK 603
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
224 -----AAVLEOEKNGKGYLPGQG-----PFVAGFASNLGDVSPNLLGPHC- 264
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
604 GNIOGLRNTKLNMAKALETTALOGNIYSDGLHVASDGHVSLANGNADFTGHHTLRKAD 663
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
265 VNTGE-----SCDNDKSTCPNGSPCMASGPGQM----- 295
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
664 VNAGSVKGRKLKADNTNITSSSG--DITLVAGNGIQLDGOKRNSINGKHISIKNGGNA 721
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
296 -----FESTH-----II 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
722 DLKNLVNAKSGALNIHSDRALSTENTKLESTHNTLHNOHERVTLNQDAYAHRHLST 781
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
303 GRITVQAKELYASASOEVTGPVLAHQVNMNTDVSVOLNATHVTCTCPALGSPAGT 362
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
782 GSOIQMNDR--LPSANKVLAVNGVLA-----LNARYS-----QIADNT 816
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
363 IDGVSGLNTGCTT--EGDPFWDLRDLQLGKPSSEIVECQKPKPLIH---SGELTI 415
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
817 TRAGAINLTAGTALVKRGNINMSTVSTLTEDNAB-----LKPGLGRINIEAGSLT 871
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
416 -PHRMPQDIVDQIVTVGSLAIAIRG-----ELTMSGRFRRAIKKEFLYGM--- 464
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
872 EPANRISANTDLSIKTGGKLLLSAKGNAGAPSAOVSLKKNIRLVGETDLRKSKIT 931
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
465 --KDMTVV-----TAGLSNVTHTTYTVE-----EYQARQEAFASTIGPHT 504
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
932 AGKNLVVATTGKLNLEAVNNSPNSVFPQKAELNQSKELQIQAQKKSS---PKS 987
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
505 -----LSAIIQLFRDLAKA-----IATDVVAMS-----SGPEPP 534
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
988 KLIPTLQEBRDLAFYIAINKEVKKKRKEYLQAKLSAQNIDISAGLEISGSDIT 1047
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
535 FPKNL-----IASLIPNIADRAPIGK-----HFGDVL--QPAKPEYRVG 571
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1048 ASKKLNLHAAGVLPKAADEMAAILIDGITDYEIEIKRPYKSHYDKAALNKPRLTGR 1107
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
572 EVVEV-----IFVGAN-----PKNSAENQTH 592
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1108 VSIHAAALDDARIITIGASEIKAPSGSIDIKAH 1140
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

RESULT 10

S75251  
hypothetical protein slr1028 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S75251

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75251

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3972 <KAN>

A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PID:BA017165.1; PID:g165  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Superfamily: *Synechocystis* hypothetical protein slr1028

Query Match  
Best Local Similarity 21.18; Pred. No. 78;  
Matches 80; Conservative 60; Mismatches 130; Indels 110; Gaps 21;

```
3.18; Score 112; DB 2; Length 3972;
105 FOYTLYLASEGFSNRTFOYIVSGIMKSIDIAHTNLRPKI---FTNKN-----VANY 155
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1645 FSYPI-----QGTENMLFLIDPG-----SSNTLIGPVEVTVYFENNDRTPFVSAPV 1693
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
156 QINRSPSSYLLNQSERARSSNTDKEMLVKLVNDGEDGLISWFAIHVPVSMNSNHF 215
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1694 QINMA-----TMGDDQIVTGTATVTEANDSLALID-----SGFI 1728
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
216 VNSDN--MGYA--AYLEOE-----KNGKGYLPGQGPVAG-----FASNL-----GD 254
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1729 INSDNPAIGVYLAATNSDCTLAVVAGNRGYSDSOGVNLNCTVOILFSCDILSGSGS 1788
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
255 VSPNLT-----GPHCVNTGESCDNDKSTCPNGSPCMASGPGQDMFESTHTIGRIIYOK 309
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1789 LSTTILNGNDGVLTINIQADGNQRL-----SLATAGD-----IDGDSIPDLVIGAP 1838
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
310 AKELYASASOEVTGPVLAH--QVNMNTDVSVOLNATHVTCTCPALGTSFAAGTIDGVS 367
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1839 MGNFAGAVYVIGSYLSNKGQIIDVTNLTSPN-----TMGFEVNGNEAEDLA 1888
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
368 GUNITGTEGDEPFMTLRDLQKPS-----EIVECQKPK---KPIILHSGE- 412
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1889 GFSVYVGNFDDGSDIVIGAPAKDSNGNRGVYLVAGFAQGSAPDSISPTVITSGKS 1948
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
413 LTRHPMPQDIVDQIVTVG 432
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
1949 FDIPIPD-QPN-PPSQTLTVG 1966
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

RESULT 11

T43369

heat-shock protein HSP60 precursor, mitochondrial - *fission yeast* (*Schizosaccharomyces*

C:Species: *Schizosaccharomyces pombe*

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000

C:Accession: T43369

R:Toshida, H.; Yanagi, H.; Yura, T.

Gene 167, 163-166, 1995

A:Title: Cloning and characterization of the mitochondrial HSP60-encoding gene of *Sch*

A:Reference number: 222456; MUID:96144268; PMID:8566770

A:Accession: T43369

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-582 <TOS>

A:Cross-references: EMBL:D50609; NID:g1229150; PID:BA009171.1; PID:g1229151

A:Experimental source: haplotype h-1eul-; strain Hml23

C:Genetics:

A:Gene: mcp60

C:Function:

A:Description: essential for cell viability at all temperatures

C:Superfamily: chaperonin groEL

C:Keywords: mitochondrion

Query Match  
Best Local Similarity 23.88; Pred. No. 5.5;  
Matches 63; Conservative 43; Mismatches 92; Indels 67; Gaps 14;

```
3.18; Score 109.5; DB 2; Length 582;
133 IDIAHTNLRGKIFINKGNVAVOINRSPSSY-----LNQSERAR-----YSSN 178
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
336 IDVSIKAPDHL-----GSGGSTVYTRKEDITIMKKGAGDHVAVDRQIRGVADNRLTE 391
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
179 TDKEMLVKLVNDGEDGLISWFAIHVPVSM--NSNHFVNSDMNGYAAYLEOEKNGKYL 237
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
392 SEKEKIQERLAKLSG--GIAVIRKVASSEVEVNEKKRIRYDALNAVAAV-----SEGYL 444
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
238 PGCG-FFVAGFASNLGDVSPN-----ILPHCVNTGESCDNDKSTCPNGSPCMASGPG 292
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```



```

Db      445 PGAGHSFVK--ASLR.LGDIPTNNFDQKLGVEIVRKA-----ITRPA 483
OY      293 QDMFESETHIIRIIYOKAKELIASAQEVTGPVLAHQWVMMDVSVOLANHTFKTKCP 352
        I : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      484 QTILEMAOLEGNLIYGLKELEYGRKEFN--IGYDIAKDREVDLNEIGV-IDPLKTVRT--- 537
OY      353 ALGVSAFGATIDVGSGLNITOGTTE 377
        I : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      538 -----GLVD-ASGVASLMGTTE 553

RESULT 12
C2HU
complement C2 precursor [validated] - human
N:Contans: Classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2a subunit; c1
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1988 #sequence revision 30-Jun-1988 #text change 23-Mar-2001
C:Accession: A25971; J50281; A25290; A05289; A37539; A37540; A26506; B26506; I56179; I56180
R:Mu. L.: Morley, B.J.; Campbell, R.D.
Cell 48, 331-342, 1987
A>Title: Cell-specific expression of the human complement protein factor B gene: evidenc
A:Reference number: A25971; MUID:87102880; PMID:3643061
A:Accession: A25971
A:Molecule type: DNA
A:Residues: 694-752 <MUL>
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59624.1; FID:g467309
R:Horiuchi, T.; Macon, K.J.; Kidd, V.J.; Volanakis, J.E.
J. Immunol. 142, 2105-2111, 1989
A>Title: cDNA cloning and expression of human complement component C2.
A:Reference number: J50281; MUID:89156483; PMID:2495504
A:Accession: J50281
A:Molecule type: mRNA
A:Residues: 1-532, 'F', 534-752 <HOR>
A:Cross-references: GB:M26301
R:Bentley, D.R.
Biochem. J. 239, 339-345, 1986
A>Title: Primary structure of human complement component C2.
A:Reference number: A25290; MUID:87127920; PMID:2949737
A:Accession: A25290
A:Molecule type: mRNA
A:Residues: 1-752 <BEN>
A:Cross-references: GB:X04481; NID:g34627; PIDN:CAA20169.1; PID:g34628
R:Bentley, D.R.; Porter, R.R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1212-1215, 1984
A>Title: Isolation of cDNA clones for human complement component C2.
A:Reference number: A05289; MUID:84144868; PMID:6199794
A:Accession: A05289
A:Molecule type: mRNA
A:Residues: 588-717 <BE2>
A:Cross-references: GB:K01236
R:Gagnon, J.
Philos. Trans. R. Soc. Lond. B Biol. Sci. 306, 301-309, 1984
A>Title: Structure and activation of complement components C7 and factor B.
A:Reference number: A37539; MUID:85038851; PMID:6149575
A:Accession: A37539
A:Molecule type: protein
A:Residues: 137-149, 'AG', 150-171; 454-466; 574-717 <GAG>
A>Note: two glycosylation sites were determined
R:Parkes, C.; Gagnon, J.; Kerr, M.A.
Biochem. J. 213, 201-209, 1983
A>Title: The reaction of iodine and thiol-blocking reagents with human complement compo
hol group.
A:Reference number: A37540; MUID:83308518; PMID:6555044
A:Accession: A37540
A:Molecule type: protein
A:Residues: 244-267, 'G', 269 <PAR>
R:Kerr, M.A.; Gagnon, J.
Biochem. J. 205, 59-67, 1982
A>Title: The purification and properties of the second component of guinea-pig compleme
A:Reference number: A26506; MUID:83022289; PMID:6922702
A:Accession: A26506
A:Molecule type: protein

```

```

A:Residues: 244-248 /'SX',251,'KK',254-256 <RES>
A:Accession: B26506
A:Molecule type: protein
A:Residues: 21-23,'X',25-28,'LUX',32-33,'S',35-38,'X',40,'X',42-44,'X',46 <KE2>
R:Ishii,Y.;Zhu,Z.B.;Macon,K.J.;Volanakis,J.E.
J.Immunol.151,170-174,1993
A>Title: Structure of the human C2 gene.
A:Reference number: 156179; MUID:93315833; PMID:8326124
A:Accession: 156179
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-532,'F',534-752 <RES>
A:Cross-references: GB:I09708; NID:g179663; PIDN:AAB97607.1; PID:g298124
R:Bentley,D.R.;Campbell,R.D.;Cross,S.J.
Immunogenetics 22, 377-390, 1985
A>Title: DNA polymorphism of the C2 locus.
A:Reference number: 154419; MUID:86032058; PMID:2997031
A:Accession: 154419
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 21-46 <RE2>
A:Cross-references: GB:M15549; NID:g187764; PIDN:AAA59649.1; PID:g187765
C:Genetics:
A:Gene: GDB:C2
A:Cross-references: GDB:119731; OMIM:217000
A:Map position: 6p21.3-6p21.3
A:Introns: 16/1; 86/1; 148/1; 206/1; 239/1; 283/3; 330/1; 377/1; 407/1; 454/1; 485/3;
C:Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C
C:Function:
A:Description: cleaves complement C3 and complement C5 alpha chains
A:Pathway: complement classical pathway
C:Superfamily: complement C2; complement factor H repeat homology; trypsin homology;
C:Keywords: complement classical pathway; duplication; glycoprotein; hydrolase; plasm
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-243/Product: complement C2b subunit #status experimental <C2B>
F:24-84/Domain: complement factor H repeat homology <FH1>
F:89-144/Domain: complement factor H repeat homology <FH2>
F:151-204/Domain: complement factor H repeat homology <FH3>
F:244-752/Product: complement C2a subunit #status experimental <C2a>
F:253-442/Domain: von Willebrand factor type A repeat homology <VFA>
F:471-739/Domain: trypsin homology #status atypical <TRY>
F:24-64,51-84,89-131,117-144,151-191,177-204,463-561,492-508,584-600,638-665,675-705/
F:249,112,290,333,657,471/Binding site: carbohydrate (Asn) (covalent) #status predicte
F:243-244/Cleavage site: Arg-Lys (complement subcomponent C1s) #status experimental
F:507,561,679/Active site: His, Asp, Ser #status predicted
F:621,651/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match          3.0%      Score 109; DB 1; Length 752;
Best Local Similarity 21.6%; Pred. No. 9;
Matches   93; Conservative 58; Mismatches 151; Indels 128; Gaps 23;

OY    262 PHCVNTGSG--CONDKSTCPNGSPSCMGASPGQDMESHTIIIGRIITYOKARELY--ASA 317
Db     133 PNGMMDEGTAVACDAGACHCPNPGLT---GAVTGRFGSH-GDKATRYKCSSMLVLTTGSS 187
OY    318 SOEVTGPVLAAHQWVNMTDVSVOLNATHPTWKTCPALGVSFPAAGTTIDVSGSLNITGGTE 377
Db     188 EREGCGKNV---WSGTPEICROPYSVDPEEDVAVALGTSFS---HMLGATNPLOKTKE 239
OY    378 GDPFWDTLRDOLLGKPSE-----ELVECKRPKRP-----TLHSGELTPPHMQPD 422
Db     240 S-----LGKKIIOIRSGHLNLYLLDDCSQSSENDPLIFKESASLMDRIFSEF 288
OY    423 I-VDOVOIVTWGSIAAIAPGELTMSGRRRRRRAIKKFALGXMKNNVVYTAGLSNYTHH 481
Db     289 INVSAVAIT-----FASEPVLKSVLNDNSR-----DMTEVISTLEN----- 325
OY    482 ITTEEYO---AORYEAASTIY-----GPHTLASAYIOFLRLAKAI--ATDTVA 525
Db     326 -ANKKHENGNGTGTYTAALSYYILMMNMQRRLGMETMA---MQEIRAIILLTDGNS 379
OY    526 NMSSGEPPPFKNIASLIPIADRADIGHFGSDVIOPAKPEYRG-----EVLEY 576

```

Db 380 NMGSGPK-----TAVDHIREFILINOKRNDYD----IYALGVGLDMDRELNEL 426  
Oy 577 IFVGANPKNSAENQ-----HQT-----LTVKEEDSVADMOIYMDASMETRFYWH 624  
Db 427 -----GSKKDEGHAFILIDTKALHOFHEHMLDVSKLFTDTCGVGMASMANASQERTPMH 481  
Oy 625 KGIILGISMNT 634  
Db 482 VTIKPKSQET 491

RESULT 13  
T05174  
hypothetical protein T6K22.50 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C:Accession: T05174  
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirke, W.; Stiekema, W.; Bancroft, I.; Men  
submitted to the Protein Sequence Database, August 1998  
A:Reference number: Z15400  
A:Accession: T05174  
A:Molecule type: DNA  
A:Residues: 1-1736 <BEV>  
A:Cross-references: EMBL:AL031187  
A:Experimental source: cultivar Columbia; BAC clone T6K22  
C:Genetics:  
A:Map position: 4  
A:Intons: 42/3; 106/3; 176/1; 304/1; 366/3; 397/2; 430/1; 484/1; 704/3; 799/2; 837/3;  
A:Note: T6K22.50

Query Match 3.0%; Score 109; DB 2; Length 1736;  
Best Local Similarity 20.7%; Pred. No. 34;  
Matches 111; Conservative 67; Mismatches 184; Indels 174; Gaps 29;

Oy 227 LFEQEKNGKYLPGQ-----GFWA-----GFAS-----NLGDVSPNIGP 262  
Db 694 LITLENKKYTL-GQSLYTPREISFTYVICTGHSNNDQITKKVIMHFSKGPYRP--LTP 750  
Oy 263 HCV-----NTGSCDNKSTCPNGSPSCMASGPGDMFESTH----- 300  
Db 751 DVQKNGIGILYVRNPGDS-----RVCEVNFPCILDMVEGSELTYITQTRSMKIKIS 806  
Oy 301 ----IIGRIYOKAKELYA-----SASQEVTPRYLAHQVNNMD--VSQNLNHTYKTKC 351  
Db 807 PKTTIGESVASVAASARGPSFSFPAITLKPDIAAPGLTLTPRIPTDEDTREFVYS-- 864  
Oy 352 PALGSFAAGTIDVSGNLITQGTTEGDFPMDTLRDQLLGKPESEIVECQPKPILLHSG 411  
Db 865 ----GTSMATPVLAGIAL----- 879  
Oy 412 ELTIPHP-WQPDIVDQIVTVGSLAIAIP-GEITTMGRRRREAIKKEFA-----LYGN 464.  
Db 880 -LKISHPMNSPAVIKSALVTT--AMKTPYGERLTVDCGNKYKVAADAFYGGGLVLEKRA 935  
Oy 465 KDMTVIAGLSNVYTHYITTYEEYQAKREASTIYG---PHTLSAYQL-----FRDL 515  
Db 936 TDPGLVYDMIDYTHYLCOSQLTYDKKVSALGNVNNKCPSSSSILDLNPSITIPDL 995  
Oy 516 AKAI-ATDVANMSSGPEPPFFKNLILASLIPNIDRAPIGKHEGDVLAQAKPEY-----R 569  
Db 996 KGTIVNTRTYTNV--GRVASYKRPV-----EAPFG--FNVVYSRKKLKFNKTRKK 1042  
Oy 570 VEEVEVITVGANPKNSAE--NOTHOTFLTVEKEDSVADMOIYMDASMETRFYWHKGI 627  
Db 1043 LAFITIVVHIGVRHDSSELVESHQRMLESFESAEARESTIVN-----YHNGF 1093  
Oy 628 LGLSNATYWHIPDPA-----YPCGIYRIYFGHNKKOKELLRAVILAEIGISSPF 677  
Db 1094 SGFAA-----RLTDSQAKQLSDRPDVFV--APNKKVE-LOSTRILYDLGLSPSP 1140

RESULT 14  
B81192

hemagglutinin/hemolysin-related protein NMB0497 [imported] - Neisseria meningitidis (C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: B81192  
R:Retlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. et al.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M. Science 287, 1809-1815, 2000  
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; PMID:20175755; PMID:10710307  
A:Accession: B81192  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1975 <TEF>  
A:Cross-references: GB:AE002406; GB:AE002098; NID:g7225720; PIDN:AAF0929.1; PID:g722  
C:Genetics:  
A:Gene: NMB0497

Query Match 3.0%; Score 109; DB 2; Length 1975;  
Best Local Similarity 18.6%; Pred. No. 42;  
Matches 129; Conservative 86; Mismatches 206; Indels 272; Gaps 31;

Oy 109 LYLASEGFSNRTFOYIVGIMKSIDIAHTNLK-PEKITNGVNAVNOINSPSYLLN 167  
Db 508 LALLADNDITAKT-----TNLTPTNLYVHGKDLNVDXLSAASH 551  
Oy 168 POSERARYSSNTDKMLVCLKVDNCGEDGLISWFAIHVSNMNSNHFNSDMGW---- 223  
Db 552 LKSDNAHITGTSKITLASK-----DMGV-----EAGSLNVTNLTNRNSGNLHQAOK 600  
Oy 224 -----AAYLEQEKNGKYLPGQ-----PVAAGASSNLGDVSPNIGPNC- 264  
Db 601 GNILRNTKLNAKALETALQGNIVSDGLHVASADGHVSLLANGNADEPTGHTLTKAD 660  
Oy 265 VNTGE-----SCDNDKSTCPNGSPSCMASGPGDM----- 295  
Db 661 VNAGSVGKRLKADNTNITSSG--DITLVAGNGILGDGKORNSINGKHSIKNGNA 718  
Oy 296 -----PESTH----- 718  
Db 719 DLKLNVAHAKSGALNHSRALSIENTKLESTHNTFLNQHHERVTLNOVDAYAHRLSLT 778  
Oy 303 GRIIYOKAKELYASAEQVTPRYLAHQVNNMDVSQNLNHTYKTKCPALGSFAAGT 362  
Db 779 GSOIWNDR--LPSANKLVANGVLA-----LNARYS-----QTADNT 813  
Oy 363 IGVSGSLNTTQGT-----EGDFPMDTLRDQLLGKPESEIVECQPKPILLH-----SGELTI 415  
Db 814 TLKAGALNLTAGTALVKKGNINNVSTVSTKTLLEDNAE-----LKPPLAGRLNIEGSGTLTI 868  
Oy 416 -PHPQPDIVDQIVTVGSLAIAIP-----ELTMMGRRRREAIKKEFALYGM----- 464  
Db 869 EPANRISATHTDLSIKTGKLLLSAKGNGAGPSQVSSLEAGNIRFLVGEIDLRSKIT 928  
Oy 465 --KDMTV-----IAGLSNVYTHYITTYE-----EYQAKREASTIYGPHF 504  
Db 929 AGKNLVAVATTKGLNLEAVNNSFSNFPQKAELNOKSKELEQOITLQKKSS---PKS 984  
Oy 505 -----LSATIQLFRLAKA-----IATDVANMS-----SGPEP 534  
Db 985 KLIPLOEBRDRLAFYIQAINEVKKKKRKEVYQAKLSAONIDISAGIEISSDIT 1044  
Oy 535 FFKNL-----IASLIPNIDRAPIGK-----HFGDVL--OPAKPEYRVG 571  
Db 1045 ASKKNLNLHAGVLPKRADESAAILIDGTTDQYEIKTPYKSHYDAALNKRSLRGLRG 1104  
Oy 572 EYEV-----IFVGAN-----PKNSAENQTH 592  
Db 1105 VSIHAAALDARITIGASEIKAPSGSIDIKAH 1137





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: July 3, 2003, 12:46:00 ; Search time 55 Seconds  
(without alignments)  
1425.912 Million cell updates/sec

Title: US-09-937-521-14  
Perfect score: 3582  
Sequence: 1 FSGYIGVGADCTGVSDI.....KPAVILAFEGISSPEVYTT 682

Scoring table:  
BLASTSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 11492915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCIT\_NEM\_PUB pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCITUS\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCITUS\_PUBCOMB pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1090.5	30.4	646	10 US-09-777-710A-1	Sequence 1, Appli
2	1090.5	30.4	670	10 US-09-777-710A-15	Sequence 15, Appl
3	879	24.5	280	10 US-09-739-907-146	Sequence 146, App
4	874	24.4	318	10 US-09-739-907-76	Sequence 76, Appl
5	131.5	3.7	1565	9 US-10-242-056-59	Sequence 59, Appl
6	127	3.5	27	10 US-09-739-907-150	Sequence 150, App
7	125	3.5	26	10 US-09-739-907-148	Sequence 148, App
8	116.5	3.3	766	10 US-09-815-242-10572	Sequence 10572, A
9	113	3.2	1981	9 US-09-928-457-38	Sequence 38, Appl
10	113	3.2	2015	9 US-10-066-551-1	Sequence 1, Appli
11	111.5	3.1	627	10 US-09-338-723A-4	Sequence 4, Appli
12	111.5	3.1	627	10 US-10-080-210-4	Sequence 4, Appli
13	106	3.0	1376	9 US-10-303-685-14	Sequence 14, Appl
14	106	3.0	1406	9 US-10-303-685-15	Sequence 15, Appl
15	104	2.9	878	9 US-10-165-049-2	Sequence 2, Appli
16	104	2.9	878	9 US-10-165-049-3	Sequence 3, Appli
17	104	2.9	878	9 US-09-905-983-48	Sequence 48, Appl
18	101.5	2.8	929	9 US-10-156-761-14135	Sequence 14135, A
19	101	2.8	929	10 US-09-815-242-12903	Sequence 12903, A

20	101	2.8	1399	10 US-09-815-242-5179	Sequence 5179, Ap
21	100	2.8	30	10 US-09-739-907-153	Sequence 153, App
22	100	2.8	613	10 US-09-996-620-10	Sequence 10, Appl
23	100	2.8	621	10 US-09-996-620-18	Sequence 18, Appl
24	100	2.8	1021	10 US-09-815-242-5471	Sequence 5471, Ap
25	100	2.8	1021	10 US-09-815-242-12544	Sequence 12544, A
26	99	2.8	1356	9 US-09-757-781-2	Sequence 2, Appli
27	98	2.7	24	10 US-09-739-907-152	Sequence 152, App
28	97.5	2.7	882	9 US-10-097-340-37	Sequence 37, Appl
29	97	2.7	527	9 US-10-128-714-3378	Sequence 3378, Ap
30	97	2.7	556	9 US-10-128-714-3561	Sequence 3561, Ap
31	97	2.7	556	9 US-10-128-714-8378	Sequence 8378, Ap
32	97	2.7	556	9 US-10-128-714-8561	Sequence 8561, Ap
33	97	2.7	1115	9 US-10-205-823-160	Sequence 160, App
34	97	2.7	1115	10 US-09-781-558-2	Sequence 2, Appli
35	96.5	2.7	439	10 US-09-374-586-2	Sequence 2, Appli
36	96.5	2.7	454	10 US-09-835-147-6	Sequence 6, Appli
37	96.5	2.7	464	10 US-09-835-147-30	Sequence 30, Appl
38	96.5	2.7	464	10 US-09-835-147-27	Sequence 27, Appl
39	96.5	2.7	473	10 US-09-835-147-29	Sequence 29, Appl
40	96.5	2.7	474	10 US-09-835-147-28	Sequence 28, Appl
41	96.5	2.7	476	10 US-09-835-147-3	Sequence 3, Appli
42	96.5	2.7	476	10 US-09-835-147-4	Sequence 4, Appli
43	96.5	2.7	478	10 US-09-835-147-8	Sequence 8, Appli
44	96.5	2.7	487	10 US-09-835-147-26	Sequence 26, Appl
45	96.5	2.7	502	9 US-10-092-063-38	Sequence 38, Appl

## ALIGNMENTS

RESULT 1  
US-09-777-710A-1  
Sequence 1, Application US/09777710A  
Patent No. US20020058305A1  
GENERAL INFORMATION:  
APPLICANT: OKINO, No. US20020058305A1om et al.  
TITLE OF INVENTION: CERAMIDASE GENE  
FILE REFERENCE: 1422-0458P  
CURRENT APPLICATION NUMBER: US/09/777,710A  
CURRENT FILING DATE: 2001-02-07  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 646  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-777-710A-1

Query Match	30.4%	Score	1090.5	DB	10	Length	646
Best Local Similarity	35.9%	Pred. No.	9e-87				
Matches	250	Conservative	118	Mismatches	254	Indels	75
						Gaps	14
QY	4	YVIGVGRADCTGVSDINLMGYKGNQVARGHILFRSRAFLADPDGSRMAFVSELC	63				
DB	5	YVFGKADITGEAEVGMGYSLEQTAGIHMQMARAFIEAAGRRIVYVNTDLG	64				
QY	64	MISQRLREVLKRLKESKYSGLYRDNVILSAIHHSAGAFQYLLYLILASGFSNRFPQ	123				
DB	65	MIFQAVHLKYLARLAKKPYGVYDENNVMLAATHTSGSGFSHYAMYNLSVIGFEKTFEN	124				
QY	124	YVSGIMKSIDIAHNNLKGKFEIKNGVAVNOIRSPSSYLLNPGSEARARSSTWDEK	183				
DB	125	ATVDDIVRSIEEAQARLPGRLEFYSGEIRNRSRLSLHKNP--DIAGEDGIDPOM	182				
QY	184	LVKLAVDNGEDDGLISWFAIHPSVMNNSNHFVSDNNGVAYVLEFQKRNKGYPGQGP	243				
DB	183	SVLSVDANGELAGALISNFPVHSTMTNANHILSPDNKGYSYHMEHDSVSR-----KSGE	237				
QY	244	VAGFSSNLSGVSPIILGPHCVNTEGSCDNKSTCPNGSPSCMAAGGQDMFEETHIG	303				
DB	238	VAAFOQTVAAGNLSPLM-----NLKPSGSGPDEFNDRTRIG	273				

```
QY 304 RIYOKAKELIYASASEVTGCVLAHQWVNMVSVOLANT--HTVKTCKPALGYSFAG 361
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 274 LROFAKAYELIYASASEVTGCVLAHQWVNMVSVOLANT--HTVKTCKPALGYSFAG 333
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 362 -TIDVSGNLITGTEGDFPMDLRLDQLGKPESEIECOCKPILHSGELTIPWQ 420
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 334 STEDGCGPLGEG---NNPFLSALGGLTGVPPELVQCAKTLIADGNKK--PYPM 369
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 421 PDIVDQIYVGSLSAIAIPGELITMGRFRRAIKERFALYGMKMTVYIAGLSNYTH 480
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 390 PVLPIOMFRIGOLELIGLAPAEFTVAGVIRRAVOAASEAGIRH--VVENGYANAYAS 447
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 481 YITTEYEOQAREASTIYGPHTLSAYIOLFPDLAKA-----IATDVA-----N 526
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 448 YITTEYEOQAREASTIYGPHTLSAYIOLFPDLAKA-----IATDVA-----N 507
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 527 MSGPEPPPEFKNLIALIPNADRAPIGKHFGDVLDPAPREYRVEVEYIFVGANKPN 586
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 508 FQGV-----VADDPYIGKSGFVLDQPPRESYIGDKVYAFVYTGHPKN-- 551
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 587 AENQHOTELTYEKY-EDSVADMOIMYNDASMETREYMHKGIILSNATYWHIPTAYP 645
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 552 -DLRTEKTELEVNIKDGKQTPVTAQYDMDMDTOYRWERVYGISASKATISIPGTEP 610
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 646 GIYRIYEGHNRKQELLKPAVILAFGISSPEFVYTT 682
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 611 GHYIRHYGNKAKNFWTK---ISEIGSTRSEVLGT 644
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

## RESULT 2

```
US-09-777-710A-15
; Sequence 15, Application US/09777710A
; Patent No. US20020058305A1
; GENERAL INFORMATION:
; APPLICANT: OKINO, No. US20020058305A1omu et al.
; TITLE OF INVENTION: CERAMIDASE GENE
; FILE REFERENCE: 1422-0458P
; CURRENT APPLICATION NUMBER: US/09/777,710A
; CURRENT FILING DATE: 2001-02-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 670
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-777-710A-15
```

## Query Match

```
Best local Similarity 30.4%; Score 1090.5; DB 10; Length 670;
Matches 250; Conservative 118; Mismatches 254; Indels 75; Gaps 14;
```

```
QY 4 YIYGVGRADCTGOVSDINLMGYGKNGONARGLTTRLSRAFIADPDGSRMAFVSVELC 63
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 29 YRFGGLKADITGGAFAVGMWYSSLEQKAGIMHMRQMARFVIEAASGRLYVNTDLG 88
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 64 MISQRLRELVKRLKESYGSLYRRDVIISAIHTHSGPAGFQYTYIILASEGFSNR 123
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 89 MFEQAHVLKRLKRYGVYDENVMYLAATHTHSGPGGFSHYAMYNLSVLCFOKRTN 148
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 124 YIVSGIMKSIDIAHTNLKPGKIFINKGNVANOINSPSSYLLNPOSEARVSSNDKEM 183
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 149 AIVDGIKSTIERAQAALQGRFLFYSGELRMSRNSLSLHLKP--DIAGYEDGIDPOM 206
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 184 LVKLVDLNGEDLGLISFPAIHPSVNNNSNHFVNSDNMGAAALFPOEKNKGYLPQCGP 243
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 207 SVLSFVDANGELAGALISWFRVHSTWNTNHLISPNKGYASVHMEHDVSR-----KSG 261
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 244 VAGFASSNLGDVSPNLTGPHCVNTGSCNDKSTCPNGGSPSCMASGPGQDMFESHITG 303
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 262 VAAFAQTNGNLSPNL-----NLKPGSGPDEFNDNTREIG 297
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 304 RIYOKAKELIYASASEVTGCVLAHQWVNMVSVOLANT--HTVKTCKPALGYSFAG 361
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

```
Db 298 LROFAKAYELIYASASEVTGCVLAHQWVNMVSVOLANT--HTVKTCKPALGYSFAG 357
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 362 -TIDVSGNLITGTEGDFPMDLRLDQLGKPESEIECOCKPILHSGELTIPWQ 420
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 358 STEDGCGPLGEG---NNPFLSALGGLTGVPPELVQCAKTLIADGNKK--PYPM 413
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 421 PDIVDQIYVGSLSAIAIPGELITMGRFRRAIKERFALYGMKMTVYIAGLSNYTH 480
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 414 PVLPIOMFRIGOLELIGLAPAEFTVAGVIRRAVOAASEAGIRH--VVENGYANAYAS 471
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 481 YITTEYEOQAREASTIYGPHTLSAYIOLFPDLAKA-----IATDVA-----N 526
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 472 YITTEYEOQAREASTIYGPHTLSAYIOLFPDLAKA-----IATDVA-----N 531
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 527 MSGPEPPPEFKNLIALIPNADRAPIGKHFGDVLDPAPREYRVEVEYIFVGANKPN 586
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 532 FQGV-----VADDPYIGKSGFVLDQPPRESYIGDKVYAFVYTGHPKN-- 575
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 587 AENQHOTELTYEKY-EDSVADMOIMYNDASMETREYMHKGIILSNATYWHIPTAYP 645
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 576 -DLRTEKTELEVNIKDGKQTPVTAQYDMDMDTOYRWERVYGISASKATISIPGTEP 634
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 646 GIYRIYEGHNRKQELLKPAVILAFGISSPEFVYTT 682
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 635 GHYIRHYGNKAKNFWTK---ISEIGSTRSEVLGT 668
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

## RESULT 3

```
US-09-739-907-146
; Sequence 146, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 36 Human Secreted Proteins
; FILE REFERENCE: P2022P1
; CURRENT APPLICATION NUMBER: US/09/739,907
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/348,457
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: 60/070,567
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,692
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,704
; PRIOR FILING DATE: 1998-01-07
; PRIOR APPLICATION NUMBER: 60/070,658
; PRIOR FILING DATE: 1998-01-07
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 146
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-739-907-146
```

```
Query Match 24.5%; Score 879; DB 10; Length 280;
Best local Similarity 81.1%; Pred. No. 8,5e-69;
Matches 163; Conservative 28; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 1 FSGYIVGRADCTGOVSDINLMGYGKNGONARGLTTRLSRAFIADPDGSRMAFVSVELC 60
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 61 FSGYIVGRADCTGOVSDINLMGYGKNGONARGLTTRLSRAFIADPDGSRMAFVSVELC 120
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 61 ELQMISQRLRELVKRLKESYGSLYRRDVIISAIHTHSGPAGFQYTYIILASEGFSNR 120
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 121 DIGMISQRLRELVKRLKESYGSLYRRDVIISAIHTHSGPAGFQYTYIILASEGFSNR 180
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 121 TFOYIVSGIMKSIDIAHTNLKPGKIFINKGNVANOINSPSSYLLNPOSEARVSSND 180
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 181 TFOHMYTGILKSIDIAHTNMKPGKIFINKGNVANOINSPSSYLLNPOSEARVSSND 240
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
QY 181 KEMLVKLVLDNGEDLGLISW 201
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

Db 241 KEMIVLKMVDLNGDGLISF 261

## RESULT 4

US-09-739-907-76  
Sequence 76, Application US/09739907  
Patent No. US20010012889A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 36 Human Secreted Proteins  
FILE REFERENCE: P2022P1  
CURRENT APPLICATION NUMBER: US/09/739,907  
CURRENT FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: 09/348,457  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: 60/070,567  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: 60/070,692  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: 60/070,704  
PRIOR FILING DATE: 1998-01-07  
PRIOR APPLICATION NUMBER: 60/070,658  
PRIOR FILING DATE: 1998-01-07  
NUMBER OF SEQ ID NOS: 196  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 76  
LENGTH: 318  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-739-907-76

Query Match 24.4%; Score 874; DB 10; Length 318;  
Best Local Similarity 80.6%; Pred. No. 2,9e-68;  
Matches 162; Conservative 28; Mismatches 11; Indels 0; Gaps 0;

QY 1 FSGYIGVGRADCTGOVSDINLMGYKNGONARGLTRFSRAFIADPDGNSNMAFVS 60  
Db 99 FSGYHIGVGRADCTGOVADINLMGYKSGNAQGLTRIXSRAFIADPDGNSNMAFVS 158  
QY 61 ELICISORLRELVKRLSKYGSIXRRDNVLSAHRHSGPAGFOFTLLYLAEGFSNR 120  
Db 159 DIGWVSQRLEVINRSLKSYGSIXRRDNVLSGTHHSGPAGFOFTVFIASEGFSNQ 218  
QY 121 TFOYVSGIMKSIDIAHTNLKPGKIFINKGNVANVOINRSPSYLLNPOSERARYSSNTD 180  
Db 219 TFOHMTGILKSIDIPHTNKKPGKIFINKGNVDSVOINRSPSYLLNPOSERARYSSNTD 278  
QY 181 KEMIVLKLVDLNGEDGLISM 201  
Db 279 KEMIVLKMVDLNGDGLISF 299

## RESULT 5

US-10-242-056-59  
Sequence 59, Application US/10242056  
Publication No. US20030113323A1  
GENERAL INFORMATION:  
APPLICANT: Ensign, Jerald C  
APPLICANT: Bowen, David J  
APPLICANT: Petell, James  
APPLICANT: Faily, Raymond  
APPLICANT: Schoonover, Sue  
APPLICANT: ffrench-Constant, Richard  
APPLICANT: Orr, Gregory L  
APPLICANT: Merlo, Donald J  
APPLICANT: Roberts, Jean L  
APPLICANT: Rochelleau, Thomas A  
TITLE OF INVENTION: Insecticidal Protein Toxins from  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: DoweIanco  
STREET: 9330 Zionsville Road

CITY: Indianapolis

STATE: IN

COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/242,056

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/063,615

FILING DATE: 18-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/395,497

FILING DATE: 28-FEB-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/007,255

FILING DATE: 06-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/608,423

FILING DATE: 28-FEB-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/705,484

FILING DATE: 28-AUG-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/743,699

FILING DATE: 06-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: Borucki, Andrea T

REGISTRATION NUMBER: 33651

REFERENCE/DOCKET NUMBER: 50301E

TELECOMMUNICATION INFORMATION:

TELEPHONE: 317-337-4846

TELEFAX: 317-337-4847

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 1565 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-10-242-056-59

Query Match 3.7%; Score 131.5; DB 9; Length 1565;

Best Local Similarity 20.4%; Pred. No. 0.047;

Matches 141; Conservative 77; Mismatches 257; Indels 215; Gaps 32;

QY 37 TRLSRAFIADPDGNSNMAFVSVELCMISGRRLLEVLKRLSKYGSIXRRDNVLSA 96  
Db 656 TYTLSEADFTDPDKN-----YLOVCL-----NVMWDHYD 685  
QY 97 THSGPAGFO-----TYLYLASGEFSNRTFOYIVS-----GIMKSIDIAHTNLKPGKI 145  
Db 686 RPSGKKGATSWVSKWFWVYVALQDSKAPDAIRLVSYDYSKRGVLYLDFTWSSL-PAKT 744  
QY 146 FINKGNV-----ANVOINRSPSYLL--NPOSERARYSSNTDKEMLVKLVDLNGEDLG 197  
Db 745 RLNTFTVTRLLEKANLGLD-SLDDYTLQADPSLEADLVTDGKSEPM-----DENGSEN-G 796  
QY 198 LISWFAIHPSVMNNSNHFVNSDNMGYAA-----YFEDE-KNKGYLPGQGFVAGFAASNL 252  
Db 797 LYEWELFPHLPFLVATRPANEOQFSPQKSLHYTFDPAKKK-----PINAP-----AY 845  
QY 253 GDVSPNITLPGHCVTGSCDNDKSTCPNGSGMCMASGPDQMFESTYIITIGRIITYOKA-- 310  
Db 846 WNVRFPLVWG-----NSDLSRHLDDSIDP-----DTQAYAH-----PVIYQKAVF 884  
QY 311 -----KELVASASQEVYGP-----VLAH 329  
Db 885 IAVVSNLIAQGDMMYROLTRDGLTQARVYVYLAELLGPRPDVSLSSIWTPOTLDTLAAG 944

QY 330 QAVNMVDVSVQINATHVTCTKPAALGYSPAGTIDGV--SGINIGCTEGDPFMDTLRD 387  
DB 945 QKAVIRDEHQLANDSDTALPALPRGNVSYLKLADNGYFNELNVLMLLS-----HMDTLDA 999  
QY 368 QLL-----CKRSEELVEGCKRPILL-----HSGLLTTPHPQPIVQVQIVTGS 433  
DB 1000 RLYNLRHNLTVDGKPLSLPLVAPVVALLAORASGTLT-----NGVSGAMLTVP 1052  
QY 434 LAIAAI-----PCELTTMSG-----RRFREAIKKEFALYGMKDMT-----VVA 472  
DB 1053 YRFSAMLPRAVSAGVTLTISFGONLSLERSERAOEELAQOQLDMSYATTLQOQALD 1112  
QY 473 GLSNVYTHYITTEEYOQORYEPAASTITGPHLSAYIQLFRDLANA-----IATDVYANNS 528  
DB 1113 GLAAORLALLASQATAQ-QRHDHYTYTL-QNNISSAEQLVMDTQTSAGSLISSSTGVQTA 1170  
QY 529 SGPEPEPFKNLNLASIPNLADRAPIGKHKGDVLQPAKPRYRGEVVEYFVGCANPRNSAE 588  
DB 1171 SG-----ALKVYIPNIEGLADGS-----RYEGYTEAIATIGMAAGQAT 1208  
QY 589 NOTHOTFLVEKEEDSVADMQIMYNDASWE 618  
DB 1209 SYVAERLATTEYRRRRREEMQIOYQQAQSE 1238

RESULT 6  
US-09-739-907-150  
; Sequence 150, Application US/09739907  
; Patent No. US20010012889A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 36 Human Secreted Proteins  
; FILE REFERENCE: P2022P1  
; CURRENT APPLICATION NUMBER: US/09/739, 907  
; CURRENT FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 09/348, 457  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: 60/070, 567  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: 60/070, 692  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: 60/070, 704  
; PRIOR FILING DATE: 1998-01-07  
; PRIOR APPLICATION NUMBER: 60/070, 658  
; PRIOR FILING DATE: 1998-01-07  
; NUMBER OF SEQ ID NOS: 196  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 150  
; LENGTH: 27  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-739-907-150

Query Match 3.5%; Score 127; DB 10; Length 27;  
Best Local Similarity 88.9%; Pred. No. 0.00019;  
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 76 RLESKYGLYRDVNLILSAITHSGPA 102  
DB 1 RLOSKYGLYRDVNLILSGTHSGPA 27  
RESULT 7  
US-09-739-907-148  
; Sequence 148, Application US/09739907  
; Patent No. US20010012889A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 36 Human Secreted Proteins  
; FILE REFERENCE: P2022P1  
; CURRENT APPLICATION NUMBER: US/09/739, 907  
; CURRENT FILING DATE: 2000-12-20

QY 10 RADCTGVSDINIMGYKNGONARGL 35  
DB 1 RADCTGVADINIMGYKSGONARGL 26  
Query Match 3.5%; Score 125; DB 10; Length 26;  
Best Local Similarity 84.6%; Pred. No. 0.00027;  
Matches 22; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 8  
US-09-815-242-10572  
; Sequence 10572, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815, 242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191, 078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206, 848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207, 727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242, 578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253, 625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257, 931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269, 308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 10572  
; LENGTH: 766  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10572

Query Match 3.3%; Score 116.5; DB 10; Length 766;  
Best Local Similarity 21.4%; Pred. No. 0.32;  
Matches 87; Conservative 55; Mismatches 139; Indels 125; Gaps 22;  
QY 331 WV-NMVDVSVQINATHVTCTKPAALGY-----SFAAGTIDVSGINIGCTEGDPFMD 383  
DB 1 WV-NMVDVSVQINATHVTCTKPAALGY-----SFAAGTIDVSGINIGCTEGDPFMD 383



```

Db 205 WISSOTDKAIAK-----DGFKIRPAKDYDNLYSALARAKADMVGLNTRALTV--KYOD 258
QY 384 TLRLDGLKPS-----EEIVECOKPKPILHS-----GELTIPHPW----- 419
Db 259 NLSGRVOTPLTANVROOEKTEIOFKPOTYTTISLTVSESEKAKMTOKNPKYALKEROEAO 318
QY 420 -----OPDIV-DVQ--ITVVGSLAIAIPGELTTM--SGRRFREAIKEFALYGMKD 466
Db 319 LKVELSKOKGLVTDIOEKVTEN-----APLPYDLEIOREANORFOFSAKTLSL----- 369
QY 467 MTVVIAGLSNRY--THYITTYEYOQOREASTIYGPHITLSAYIOLFPDLAKAATOTVA 525
Db 370 -----VQSLYETHKIVSY-----PRTDISKY-----LTTDTKG 396
QY 526 NMSS-----GPE-PPEFKNLASLIPNIADRAPICKHFGDVLQPAKFEY----- 568
Db 397 TMKRLALADFSPEVAGYLKNGAVVQOQKAFQNAKYTDHGLIPTERFRFEKLSNDEQ 456
QY 569 RVGEVEVEIVG--ANPRNSAENQTHQTLTVEKYEEDSVADQIMYNDASMET----- 619
Db 457 KIYQMIYORFLGLFAEP-----NOTKQTKVAFGKETFEVHQNKVVAVAGKMTAEQPLS 511
QY 620 RFYHHKGLIGLSNATYIWHIPDTAVPGIYRIRFGHNRKQELLAPA 665
Db 512 TVQOKGMVAPNFTINKELTSPKP--LTEGTLGKMEKHSIGTPA 556

RESULT 9
US-09-928-457-38
; Sequence 38, Application US/09928457
; Patent No. US20020164603A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA, specific proteins and peptides
; TITLE OF INVENTION: of the Neisseria meningitidis species bacteria, method
; NUMBER OF INVENTION: for obtaining them and their biological application.
; NUMBER OF SEQUENCES: 99
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30 (OEB)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/928,457
; FILING DATE: 2001-08-14
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/214,759
; FILING DATE: 199-12-10
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1981 acids amin,s
; TYPE: acide amin,
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..1981
; US-09-928-457-38

Query Match 3.2%; Score 113; DB 9; Length 1901;
Best Local Similarity 18.9%; Pred. No. 2.9;
Matches 131; Conservative 86; Mismatches 204; Indels 272; Gaps 32;

```

```

Db 604 GNIOLRNTKLAAKALETTALOGNIVSGLHVAADGSHVSLANGNADFTGHNTLTAKAD 663
QY 265 VNTGE-----SCNDKSTCPNGSGSMKMAAGPGQDM----- 295
Db 664 VNAGSVGKRLKADNTNITSSG--DITLVGNGIQLDDGKORSINSKHIKNGGNA 721
QY 296 -----FESTH----- 302
Db 722 LKNNLNVAKSGALNIHSDRALSTENTKLESTHNTHLNAQHERVTLNVDYVAHRHLSIT 781
QY 303 GRITVOKAKELVYASQEVTPVLAHQWVNMVTVSVOLNTHVYTKCPALGYSPAAGT 362
Db 782 GSOIWOONDK--LPSANKLVANGVLA-----LNARYS-----QIADNT 816
QY 363 IDVSGNLITOGTT--BGDPFWTLRLDQGLKSPSEELVECOKPKPILH-----SGELTI 415
Db 817 TLKAGAILMTAGTALVKKGNINMSTVSTKLEDAE-----LKLPLAGLNIENAGSGTLLTI 871
QY 416 -PHEQPDIVQVITVGSIAIAIPG-----ELTMSGRFREAIKKEFALYGM-- 464
Db 872 EPANRISAHDTLSTIKTGKLLLSAKGNAGAPSAQVSSLEAKGNIRLVTTGETDLRGSKIT 931
QY 465 --KDMTVV-----IAGLSNVYTHYITTYE-----EYQAOQRYEASTIYGPH 504
Db 932 AGKNLVAVTGTGKLINIEAVNNSFSNVPYFOKAELNCKSELEQOIAQLKSSS-----PKS 987
QY 505 -----LSATYDLFRDLAKA-----IADTYANMS-----SGRPP 534
Db 988 KLIPTEERDLRAFYQAIINKEVKKPKKEVYLOAKLSQNDILISAQIEISGSDIT 1047
QY 535 FFKNL-----IASLIPNIADRAPICK--HFGDVL--OPANPEYRWG 571
Db 1048 ASKTLNLHAAGVLPKRAADSEAAIILDSITTOYELEGKTYKSHDKALNKRPSRLTGTG 1107
QY 572 EVEVEY-----IFVGAN-----PKNSAENQTH 592
Db 1108 VSIHAAALDDARIIIGASEIKAPSGSIDIKAH 1140

RESULT 10
US-10-066-551-1
; Sequence 1, Application US/10066551
; Publication No. US20030100071A1
; GENERAL INFORMATION:
; APPLICANT: Apicella, M. A.
; APPLICANT: Gibson, J. L.
; APPLICANT: Gibson, B. W.
; APPLICANT: Scheffler, K.
; APPLICANT: Brown, E.
; TITLE OF INVENTION: Vaccine and compositions for the
; TITLE OF INVENTION: Prevention and treatment of Neisserial infections
; FILE REFERENCE: 8/75,045US1
; CURRENT APPLICATION NUMBER: US/10/066,551
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/344,452
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/310,356
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/266,070
; PRIOR FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 1
; LENGTH: 2015
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
; US-10-066-551-1

Query Match 3.2%; Score 113; DB 9; Length 2015;
Best Local Similarity 18.9%; Pred. No. 2.9;
Matches 131; Conservative 86; Mismatches 204; Indels 272; Gaps 32;

```

```

Db      511 LALLADNDITAKT-----TNLTPNENLYHTGKDLNLVNDKDLASASIH 554
OY      168 POSERARSSNTDKEMLVKLVLDNGEDGLISMFAIHPSVMNSNHFFVSDMWG-----223
Db      555 LKSDNAHITGTSKITLASK--DM-GVEAGCL-----NVTNTLRTNSGMLHIOAK 603
OY      224 -----AAVLEFQEKNNKGYLPGQ-----PVAAGFASNLGDVSPNIIQPHC- 264
Db      604 GNQLRNTKLNAKALETTALQGNIVSDGLHVASADGHVSLLANGNADFGHNTLRAKAD 663
OY      265 VNTGE-----SCDNKSTCPNGSPSCMASGPGOM-----295
Db      664 VVAGSVGKRLADNTNITSSG--DITLVAGNGIGLDGOKRNSINGKHISIKNNGMA 721
OY      296 -----FESTH-----II 302
Db      722 DLKLNVAHAKSGALNHSDBALSTIENTKLESTHNTLNAOHEVTLNOVDATAHRLST 781
OY      303 GRIYOKAKELYASASQEVTPYLAHOMVMTDVSVOLNATHVTCTCKPALGYSFAGT 362
Db      782 GSQIMQNDK--LPSAKKLVANGYLA-----LNARYS-----QIADNT 816
OY      363 IGVSGSLNTGQTT---EGDPFMDTLRDQLGKPSSEIVECOKPKPILH---SGELT 415
Db      817 TLKAGINLTAGTALVKRCNINMTSTVSTKLEEDNAE---LKPLAGRLNIEAGSGTLT 871
OY      416 -PHMPQDVLVDQIVTVGSLATAIAPG-----ELTMSGRFRRAIKKEFALYGM---464
Db      872 EPANRISANTDLSTKTKGKLLLSAKGNAGPSAOSVLEAKGNITLVGETDLKRSKIT 931
OY      465 --KDMTVV-----IAGLSNVTYHTTYE-----EYQARYEASTIIGPHT 504
Db      932 ACKNLVVATTGKLNIEAVNNSPSNYFPQKAELNQGSKLEQJLAQJLKS--PKS 987
OY      505 -----LSATIQFLRLAKA-----IATDVANMS-----SGPEPP 534
Db      988 KLIPLQERDRILAFYIOAINKEVKKKPKKREYIOAKLSAQNIDILISAGLEISGSDT 1047
OY      535 PFKNL-----IASLIPNADRAPICK-----HFGDVL--QPAKPEYRVG 571
Db      1048 ASKKLNLHAAGVLPKAADEMAAILIDGTDQYEIGKPYTKSHYDAALNKPRLGRTG 1107
OY      572 EYVEV-----IEVGAN-----PKNSAENQTH 592
Db      1108 VSIHAAALDDARIIIGASEIKAPSGSIDIKAH 1140

RESULT 11
US-09-338-723A-4
; Sequence 4, Application US/09338723A
; Patent No. US20020019038A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Wang
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC361-2
; CURRENT APPLICATION NUMBER: US/09/338,723A
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 09/220,871
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 627
; TYPE: PRF
; ORGANISM: Bipolaris spicifera
US-09-338-723A-4

Query Match
Best Local Similarity 20.8%; Pred. No. 0.63;
Matches 129; Conservative 76; Mismatches 230; Indels 185; Gaps 35;
145 IFKNGVNAVNOINSPSSYLLN-POSERARYS--NTDKEMLVKLVLDNGEDGLI-- 199

```

```

Db      13 VSIAGK-IYGVALSERPAKFVNDTPDEKKAALASIVEDPADVNNLKMOSPEYBLIR 71
OY      200 SFAIRIPVSMNN--SNHFVNSDMWGYAATL--FEQEKNNKGYLPGQGPFAVGSNLGV 255
Db      72 QPLPPIPAKEPNKLTIPVTKELIYETIKPFTQVYLSLRAR--LVGIGISPG-- 126
OY      256 SPNIIQ-----HCVNTGESCDNDKSTCPNGSPSCMASGPGDMFESTHIIQRIYQ 308
Db      127 -PTIIVPRGTEAVVRING--DRESSIHLHGPSRAFPDGDADM-----IMNG 173
OY      309 KAKELYASASQEVTPYLAHOMVMTDVSVOLNATHYKTKCPALGYSFAGTIDVSG 368
Db      174 EYKDYYPNNQ-----AARFLWYHD-----HAMHVT-----AENAYPGQAG 209
OY      369 LNTGQTTGDDPMDTLRQQLGKPS---EEIVECOKPKPILHSGELIIRHPQDVL 424
Db      210 AVLIT-----DEPEDA-----LGLPBGYGYKD-----PLVLSKYYNADGTLKTSVG 252
OY      425 DVOIVTVGSLATAIAPGELTMSGRFR-----EAIKKEFALYGMK--DMTV-----VIA 472
Db      253 EDSKSVGDIILHVNGQWPFLNVEPRKYLRLFLNAASRNFALYFKQDNTATRLPROVIA 312
OY      473 GLSNVTYHTTYEY--QAQRYEASTI--YGPHTLSAYIOLFRDLAKA-----518
Db      313 SDAGLLTHPQTSMDVAAAEYIEVDEPAPYAGTLD-----LRNFAKANGIGTDDYA 367
OY      519 -----IATDVANMSGSGPPFFKNLILIPNADRAPICKHFDVLOPAPEYR 569
Db      368 NTDKVRHFHSSQTVVDSNVPEQ-----LSQIOFPADKTDIDHFR--FHRINGEWR 418
OY      570 VGEV-----VEVIEVGANPK-----NSAENQTHQ-----TELTVEKEDSVADMOI 610
Db      419 INIGRADVENVRLAVPRTGYVELMELENSGGSWSPHIVHLVDFVVARYGEGRTGYM 478
OY      611 M-----NDASWETRYMKRGILGSLNATYIWHIPTAIRGTY-----648
Db      479 PFEAGLKVVWVGR--HETVL-----VEAHY--APWDGYMFCHNLIHEDODMAA 527
OY      649 ----RIRYFGHNRKQELLKP 664
Db      528 FDTKRLQNGCYNETTDFHDP 547

RESULT 12
US-10-080-210-4
; Sequence 4, Application US/10080210
; Patent No. US20020142423A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Huang
; TITLE OF INVENTION: Phenol Oxidizing Enzymes
; FILE REFERENCE: GC561-3
; CURRENT APPLICATION NUMBER: US/10/080,210
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: US 09/220,871
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 09/338,723
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 627
; TYPE: PRF
; ORGANISM: Bipolaris spicifera
US-10-080-210-4

Query Match
Best Local Similarity 20.8%; Pred. No. 0.63;
Matches 129; Conservative 76; Mismatches 230; Indels 185; Gaps 35;
145 IFKNGVNAVNOINSPSSYLLN-POSERARYS--NTDKEMLVKLVLDNGEDGLI-- 199

```

Db 13 VSIANG-ITGVALSRAKPVNDTPDEEKALASIVEDBADVNNMLKMCSPYPLIER 71  
OY 200 SWEAIPHYSMNN--SNHFVNSDNMGYAYL--PFOEKNKGYLPQGGFFVAGFASSNLGDV 235  
Db 72 OPLFIPPAKEPKLNPFYTNKEIWIYELVIFPFOQYVPSLRPAR--LVGYDISPG-- 126  
OY 256 SNNLGP-----HCVNTGESCNDKSTCPNGSPKMGASCPRODMFESHIIIGRIIYO 308  
Db 127 -PITIVPREGTEAVNRING--DRESSIHLGSPSRAPFDGMADM-----IMKG 173  
OY 309 KAKELYASASOEVTGPVLAHOMVNMTOVSQVNLNTHVTKCPALGYSFNAGITDYSV 368  
Db 174 EYKDYYPNNQ-----AARFLMYHD-----HAMHY-----AENAYFGQAG 209  
OY 369 LNTGCTEGDPFMDLTDOLLGKPS--EETVECKPKPILLHSGELTIPHPQPDIV 424  
Db 210 AYLT-----DPAEDA-----LGLPSGCKYDI-----PLVLSKYYNADGTIKTSVG 252  
OY 425 DVOIVTVGSLAIAIPGELTTMGRFR-----EAIKKEFALYGMK-DMTV-----VIA 472  
Db 253 EDKSVMGDILHVNQCPWPLNVEPRKRLFLNNAVSRNFALYFVQODMTATRLPFOVIA 312  
OY 473 GLSNVYHYITTYEY--QAORYEASTI--YGPHTLSAYIQLFRLAK----- 518  
Db 313 SDAGLTHPVOTSDMYAAAEERYEIVDFAPYAGQILD-----LRNFAKANGIGTDDYA 367  
OY 519 -----IATDVANSSGPEPPFKNLIASLIPRIADRAPIGKFGIIVLOPARREY 569  
Db 368 NTDKVRRFHVSSQTVVDSVYVEQ-----LSQIOPADKTDIDHFR--FRTNGEMR 418  
OY 570 VGEV--VEVIFVGANKP-----NSAENOTHO-----TFLVEKYEDSVAMQI 610  
Db 419 INGIGFADVENKVLAKVRCIYELMELNSSGMSHPHVLVDFRVARAYGDEGTGVM 478  
OY 611 MY-----NDASWETRPYMKGLIGLSNATYIWHIPDAYPGY----- 648  
Db 479 PYEAGLKVYVWIGR--HETVL-----VEAHY--APWDGVYMHCHNLIHEDQDMAA 527  
OY 649 -----RTIRYGNHKKOELLKP 664  
Db 528 PDVTKLONFGNETTDFHDP 547

RESULT 13  
US-10-303-685-14  
; Sequence 14, Application US/10303685  
; Publication No. US20030100005A1  
; GENERAL INFORMATION:  
; APPLICANT: Exelixis, Inc.  
; TITLE OF INVENTION: CRBS AS MODIFIERS OF BRANCHING MORPHOGENESIS AND METHODS OF USE  
; FILE REFERENCE: EX02-125C  
; CURRENT APPLICATION NUMBER: US/10/303,685  
; CURRENT FILING DATE: 2002-11-25  
; PRIOR APPLICATION NUMBER: 60/333,388  
; PRIOR FILING DATE: 2001-11-26  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 1376  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-303-685-14

Query Match 3.0%; Score 106; DB 9; Length 1376;  
Best Local Similarity 22.7%; Pred. No. 6.6;  
Matches 77; Conservative 41; Mismatches 97; Indels 124; Gaps 22;

OY 73 VLKRESKYSILRRDNTLSAITHSGPAGFOYTLVLAEGESNRTFOYISGIMKS 132  
Db 725 VIFLDESYSY-----PTISLSMVRTLQPSGL-----LLALE--NSTYQYI----- 763  
OY 133 IDIAHNTLKPCKIFINKGVANVQINRSPSSYLLNPOSERARYSSNTDKEMVLYKVDIN 192

Db 764 -----RWVLEGRILA-----MLTPNSPK-----LVYKFV--LN 789  
OY 193 GEDIGLSWFAIHPVSNMNSHFPVNSDNMGY--AAVLEPCKRNK-----GYLPQO-----G 241  
Db 790 DGNVHLIS-LKIKPKYI--ELYOSSQNLGFIASSTWIKIEKGIVYIGGLPKQETELNG 845  
OY 242 PVVAG-----FASNLGDVSPN-----ILGPHCVNTGESCNDKS-----TCPNGGPRSMC 286  
Db 846 GFFKGCIDVRLNNQNL--EFPNPTNNASLNPVLYVNTQCCAGDSCSKSPCHNG-- 900  
OY 287 MASGPGDMFESTHIIIGRIIYOKAKELYASASOEVTGPVLAHOMVNMTOVSQVNLNTH 346  
Db 901 -----VCHSRWDDFSCCPALTSKACAEYQWCGFSF-----CPHG 936  
OY 347 VTKCKPAL--GYSFAGTI--DGVSGL-----NITGCTT 376  
Db 937 AQ--CQPVLOGFECIANAVFNQSGQILFRSNGNITRELT 974

RESULT 14  
US-10-303-685-15  
; Sequence 15, Application US/10303685  
; Publication No. US20030100005A1  
; GENERAL INFORMATION:  
; APPLICANT: Exelixis, Inc.  
; TITLE OF INVENTION: CRBS AS MODIFIERS OF BRANCHING MORPHOGENESIS AND METHODS OF USE  
; FILE REFERENCE: EX02-125C  
; CURRENT APPLICATION NUMBER: US/10/303,685  
; CURRENT FILING DATE: 2002-11-25  
; PRIOR APPLICATION NUMBER: 60/333,388  
; PRIOR FILING DATE: 2001-11-26  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 1406  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-303-685-15

Query Match 3.0%; Score 106; DB 9; Length 1406;  
Best Local Similarity 22.7%; Pred. No. 6.9;  
Matches 77; Conservative 41; Mismatches 97; Indels 124; Gaps 22;

OY 73 VLKRESKYSILRRDNTLSAITHSGPAGFOYTLVLAEGESNRTFOYISGIMKS 132  
Db 725 VIFLDESYSY-----DTISLSMVRTLQPSGL-----LLALE--NSTYQYI----- 763  
OY 133 IDIAHNTLKPCKIFINKGVANVQINRSPSSYLLNPOSERARYSSNTDKEMVLYKVDIN 192  
Db 764 -----RWVLEGRILA-----MLTPNSPK-----LVYKFV--LN 789  
OY 193 GEDIGLSWFAIHPVSNMNSHFPVNSDNMGY--AAVLEPCKRNK-----GYLPQO-----G 241  
Db 790 DGNVHLIS-LKIKPKYI--ELYOSSQNLGFIASSTWIKIEKGIVYIGGLPKQETELNG 845  
OY 242 PVVAG-----FASNLGDVSPN-----ILGPHCVNTGESCNDKS-----TCPNGGPRSMC 286  
Db 846 GFFKGCIDVRLNNQNL--EFPNPTNNASLNPVLYVNTQCCAGDSCSKSPCHNG-- 900  
OY 287 MASGPGDMFESTHIIIGRIIYOKAKELYASASOEVTGPVLAHOMVNMTOVSQVNLNTH 346  
Db 901 -----VCHSRWDDFSCCPALTSKACAEYQWCGFSF-----CPHG 936  
OY 347 VTKCKPAL--GYSFAGTI--DGVSGL-----NITGCTT 376  
Db 937 AQ--CQPVLOGFECIANAVFNQSGQILFRSNGNITRELT 974

RESULT 15  
US-10-165-049-2  
; Sequence 2, Application US/10165049  
; Publication No. US20020192724A1  
; GENERAL INFORMATION:

APPLICANT: Brenner, Michael  
APPLICANT: Cepek, Karyn  
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin  
FILE REFERENCE: L00560/70010DERP  
CURRENT FILING DATE: 2002-06-07  
PRIOR APPLICATION NUMBER: USSN 08/237,919  
PRIOR FILING DATE: 1994-05-03  
PRIOR APPLICATION NUMBER: PCT/US 95/05518  
PRIOR FILING DATE: 1995-05-03  
PRIOR APPLICATION NUMBER: USSN 08/732,429  
PRIOR FILING DATE: 1996-11-01  
PRIOR APPLICATION NUMBER: USSN 09/798,267  
PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 878  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: (800)..(808)  
OTHER INFORMATION: HAV tripeptide  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: (2225)..(2295)  
OTHER INFORMATION: transmembrane  
US-10-165-049-2

Query Match 2.9% Score 104; DB 9; Length 878;  
Best Local Similarity 18.6% Pred. No. 4.9;

Matches 130; Conservative 94; Mismatches 255; Indels 220; Gaps 32;

QY 47 ADPGSNRAEYVSVELCMISORLELV-----LKR-----LESYIG 82  
DB 41 APPEKRPRLGRVNFEDCTGRORTAIFLPKVGTDGVITVPRPLRPHNPTDPFLGLRMD 100  
QY 83 SLYRR--DNVILSAIHTHSGPAGFFQYLYLIASGFSNRPFOYIVSGIMKSIDIAHTNL 140  
DB 101 STYRFEFTKVTLNTVGHHRPP-----PHQASVSGIOAEL-LTFPNS 141  
QY 141 KPGKIFINKGNVANYQINKSPSSYLLN---POSERARY-----SSNTDKEMLVKLV 189  
DB 142 SPG-----LRRQKRDWVLPISCPENEGKPEPKLVQIKSNKDEKGVFSYI 188  
QY 190 DLNGEDLGLISWFAI-----HPVSMNNSNHFVNSDMGYA 224  
DB 189 TGQGADTPPVGVFIIRETGWLKVTPEPLDRERIATYTLFSAVSSNGN---AVEDPMETL 245  
QY 225 AYLFPOEKNK-----GYLPQGPFFVAGFASNLGDVSPNLPCHVCNTGES 270  
DB 246 ITVTQDNQNKPEFTQEVKRGSMEGALPC-----TSMEV-----TATD 284  
QY 271 CDNDKSTCPNGSPSCMASG---PGQDMF---ESTHIIIGRIIYQAKELVASASQEVGT 323  
DB 285 ADDDVNTY-NAIAIATYILISOPELDPKMTINRNTGVISVTTGLDRSFPTYLTVQA 343  
QY 324 PYLAHQVNMVDVSYQNLATHTVTKCKALGYSPAAGTIDG-----VSGLINTQ 373  
DB 344 ADLOG-EGISTATAV---ITVTDNDNPI---FNPTTYKGVDPENANVVITTLKVPD 396  
QY 374 GTTEGDPFMDTLRDOLGKPSSEIEVCOQR-----KPLLHSGELTIPHPQOPDI 423  
DB 397 ADAPNTPAMEAVY-TIILNDGGOFYVTTNPVNDGILKTAKGLDEPAKQOYILHVAVTNV 455  
QY 424 VDOQIVTVGSLAIAIPGELTTMSGRRF-----REAIKKEFALYGMKDMTVYIAGLSNVY 478  
DB 456 VPEFVSLTSTATVTV-DVLVDNNEGIFVPEPEKRVSEDFGV--GQELITSYTAQEPDTE 512  
QY 479 THYITTYEYQAKREASTIIGPHT--LSAVIQLFRD-----LAKAIATDPVAN 526

DB 513 MEOKITVRIMWDTNRNMLE---INPDGAISTRADLREDEFEHVNKSTYTAIIITDNGSP 569  
QY 527 MSSGPEPFFKNLIASLIPNIADRAPIGKHGVDLQPAKPEYRVGEVVEVIFVGANKR-- 584  
DB 570 VATG-----TGTILLIISDVNDNAPI-----PEPR-----IFFCERNKRPQ 606  
QY 585 ---NSAENQTHQTFIVKEYEDSVADQIYNDAWET 619  
DB 607 VINIHADLPPNTPSPFTALETHGRVPMTIQYNDPTQES 645

Search completed: July 3, 2003, 12:56:14  
Job time : 57 secs